

Neuroethics needed

Researchers should speak out on claims made on behalf of their science.

How would you feel if you had to tell the truth, and nothing but the truth, for a day? Did your wife really look so good in that dress? Were you honestly late just because the traffic was bad? Did you actually do all the necessary controls for that experiment?

Society would be a different place if all our lies, however trivial, were abandoned in favour of blunt honesty. In some areas, such as criminal prosecution, this might be advantageous. In others, where little white lies help life run smoothly, knowing all the facts might be uncomfortable.

These thoughts are brought to the fore by the arrival of two US start-up companies, No Lie MRI and Cephos, which are about to offer functional magnetic resonance imaging (fMRI) brain scans in order to detect lies. The companies, which plan to launch their services later this year, say their goal is to help exonerate the innocent, and to replace the widely discredited polygraph machines used by US government agencies for screening their staff (see page 918).

Many neuroscientists think the claims being made for fMRI are overblown. They warn that there is scant evidence that it can reliably distinguish a lie from the truth in any individual case, especially in the real-life, high-stakes situations in which it might be applied.

Ethicists worry even more about what would happen if one day the scanning technique could be used to accurately discern people's inner secrets. Society would, for the first time, hold in its hands a reliable tool with which to finger deceit, and this could have a profound impact on individual privacy and human rights.

It is too early to tell if fMRI will ever be able to pinpoint liars in anything but a few, tightly controlled circumstances. Studies so far have been conducted under such conditions, and do not reflect the many types of lies and the situations they may be used to investigate. Some will argue that the mere threat of an accurate lie test could be used to extract valuable information — the same argument, in essence, that led to the use of polygraphs in the United States.

The question of how far such approaches should be taken is just one of a number of pressing ethical issues raised by the rapid recent

progress of neuroscience. So it is appropriate that last month, a group of prominent scientists, ethicists and lawyers gathered at the Asilomar conference centre in California to found the Neuroethics Society, which will address these issues.

This effort is to be applauded, but there is a lot of work to do before it can engage a rapidly expanding neuroscience community that has been relatively slow to recognize its own responsibility to address potential abuses of knowledge.

Geneticists held their own landmark Asilomar meeting more than thirty years ago to discuss the possible regulation of recombinant DNA. Today, the ethical and social repercussions of genetics are a standard component of undergraduate education. But ethics is a long way from attaining corresponding status within neuroscience. One prominent bioethicist reports that his own lecture on neuroethics was cancelled, falling victim to timetable pressure.

Neuroscientists have reasons for their reluctance to wade into ethics. The questions raised are likely to be open-ended, and their arrival in the world outside the laboratory may be some way off. Whereas a genetic test can say something definitive about a particular genetic make-up, and therefore about predisposition to disease, for example, an fMRI scan is just an indirect

measure of neural activity based on oxygenated blood flow. For now, neuroscientists have only the most basic grasp of what this says about how the brain processes information.

Even so, the arrival of No Lie MRI and Cephos suggests that fMRI is entering the 'real world', whether neuroscientists consider it ready or not. The community needs to broadcast its doubts about this situation from the rooftops — and prepare for a prolonged, complex and occasionally frustrating engagement with the public on the ethical ramifications of its work. ■

"The arrival of No Lie MRI and Cephos suggests that fMRI is entering the 'real world', whether neuroscientists consider it ready or not."

Urgent but balanced

Energy problems demand a coherent solution, not a quick fix.

The energy issues facing many of the world's governments are now acute. And there is a disturbing tendency for this urgency to generate polarizing debates on plans that could have only a marginal effect on the unfolding crisis. In the United States, such an argument has taken place over oil drilling in the Arctic National Wildlife Refuge. In Britain, an almighty row is looming over the replacement of a small number of ageing nuclear power stations.

The urgency of the current crisis is driven by stubbornly high oil prices, the clear need to do something about greenhouse-gas emissions, and the benign neglect that has characterized many national energy policies for the past two decades. But the crisis demands more than the flailing efforts of governments or political parties to develop headline-grabbing initiatives. It calls, instead, for a thorough, rational and rapid analysis of how effective energy policies should be rebuilt.

A report released this week by the Royal Society of Edinburgh (RSE) attempts to provide such an analysis for Scotland — a small country whose energy issues are not untypical of those facing Western Europe.

The exhaustion of North Sea oil and gas, together with the rapid



ageing of coal and nuclear power stations, presents a challenge for the Scottish Executive and the British government in London. Under Scotland's 1999 devolutionary settlement, London sets the energy policy, but Edinburgh is responsible for implementing it.

As the RSE's report explains, the energy crisis won't be addressed either by building wind-farms on the Isle of Lewis, or by licensing replacements for the nuclear plants that now produce half of Scotland's electricity. What's needed instead is a wide diversity of approaches to electricity generation and to energy use and distribution, and a comprehensive, integrated strategy for their implementation. The report's 37 recommendations are not a cop-out from making choices, but are instead a realistic acknowledgement of the breadth of the problem.

Some of the recommendations concern such humdrum matters as making more effective use of existing planning regulations to construct energy-efficient buildings — an area in which Britain has always lagged behind its neighbours. The report also calls for a more diverse approach to renewable energy, pointing out that the Scottish Executive's existing incentive system serves to promote wind capacity (which can be built now) over alternatives such as wave power that require further research and development. And it suggests that the executive should promote research collaborations in technologies such as clean coal and wave power, in which local

industrial companies and researchers have international expertise.

But the most important recommendation is perhaps the dullest: a call for an agency, outside the government but answerable to it, that will formulate, champion and implement an energy strategy. This demand is also pertinent to Britain as a whole, which shut its energy department in 1992. But it isn't at all clear whether the resources will be provided to establish such an agency.

Maxwell Irvine, the physicist who chaired the RSE panel, fears that without a comprehensive energy strategy, Britain's privately owned electricity industry will repeat the 'dash-for-gas' — the Thatcherite energy policy of the 1980s. That policy caused the nation to burn its irreplaceable North Sea gas supplies, rendering itself dependent, in the long term, on imported gas to keep the lights on.

The RSE committee, which raised funds for the study from foundations and other sources, is to be congratulated on bringing some sanity to an energy debate that is becoming unhinged from reality. In elections next spring, for example, the ruling coalition in the Scottish parliament looks set to split over the issue of nuclear power.

What is needed instead, in Scotland, the United Kingdom, the United States and elsewhere, is a rational energy strategy, executed by a competent agency. It is only this that can take energy usage over many years to a sustainable position, with no more lurches from complacency to crisis. ■

The mad technologist

Hollywood warms to science, but fears technology.

Cinema has been around for more than 100 years now, but the world has not tired of it. According to the Motion Picture Association of America, 9.6 billion tickets to the cinema were sold around the world in 2004. Film was, arguably, the single most important artistic medium of the twentieth century. Its resonance and power alone make it an object worthy of study.

It is also an art that technology has rendered possible. More than an opera, a play or a novel, film is a technological product, spawned directly by the inventions of electricity and celluloid. Before the Lumière brothers' *cinématographe* showed moving images of Lumière factory workers knocking off work for the day, the groundwork was laid by the invention of the pinhole camera and Eadweard Muybridge's motion-capturing photographs, for example.

So how does this most modern of media see science? On-screen, apart from a few earnest biopics of Louis Pasteur or the Curies, scientists are often portrayed as comically inept eccentrics or evil geniuses bent on world domination. They frequently re-enact Frankenstein-creator Mary Shelley's lesson about playing god — and are almost as frequently dispatched by their own twisted creations.

Yet scientists are less interested in creating things than in finding things out — discovering new species, rather than manufacturing them. Many 'mad scientists' on film are really engineers of one kind or another. They are technologists.

This muddling of science and technology is common, and it could be argued that scientists haven't helped things by claiming full credit for technologies — such as nuclear power or modern medicine —

that they merely made possible through their discoveries. If we tease science and technology apart, we find that pure scientists are often treated kindly by film-makers, who have portrayed them sympathetically, as brooding mathematicians (*A Beautiful Mind*) and heroic archaeologists (*Raiders of the Lost Ark*).

It is technology that movie-makers seem to fear. Even the best-loved science-fiction films have a distinctly ambivalent take on it. *Blade Runner* features a genetic designer without empathy for his creations, who end up killing him. In *2001: A Space Odyssey*, computers turn against humans, and *Star Wars* has us rooting for the side that relies on spiritual power over that which prefers technology, exemplified by the Death Star. Why would an inherently technological medium seem to be so wary of its own creator?

Modern technology has changed lives considerably, as becomes clear when we consider that television has been around for only about 70 years. No doubt technology has improved lives in ways that can easily be felt and measured. But it takes some getting used to. Many people find the constant introduction of new gadgets and the faster pace of life alienating and exhausting. It may be that our first completely technological art form is the one best suited to exploring how we feel about modern life.

At any rate, the mad scientist is not the only image of science. Films such as *Primer* and *Schläfer* (see page 922) have given the world a more realistic look at scientists, albeit ones with rather more interesting lives than most. There is drama and pathos to be found in the laboratory, even without the wild hair and insatiable desire to rule the world. ■

"Films such as *Primer* and *Schläfer* have given the world a more realistic look at scientists, albeit ones with more interesting lives than most."

RESEARCH HIGHLIGHTS

CANCER BIOLOGY

Skeleton key

Proc. Natl Acad. Sci. USA **103**, 9643–9648 (2006)
Researchers in France have identified a promising new target for treating patients with breast cancer that has spread to their bones.

When cancer cells settle in bone they secrete signalling molecules known as cytokines that trigger the bone's destruction, so helping the tumour to grow. Olivier Peyruchaud from the INSERM U664 institute at the University of Lyon and his colleagues have previously found that lysophosphatidic acid (LPA), a lipid produced by blood platelets, stimulates the production of cytokines.

Now they show that LPA's action is mediated through one specific receptor, LPA₁, on the surface of the cancer cells. Blocking this receptor reduced the progression of bone metastases in mice.

NEUROSCIENCE

Look who's thinking

Nature Neurosci. doi:10.1038/nn1729 (2006)

Researchers in Sweden suggest that mirror neurons — which fire both when an action is performed and when it is observed — may underlie the development of certain thinking skills in infants.

Terje Falck-Ytter of Uppsala University and his colleagues studied how well adults, 12-month- and 6-month-old babies were able to predict the outcome of a simple human action. They showed the subjects a video of a hand transferring objects into a box. Adults and 12-month-olds predicted the outcome by directing their gaze at the box; 6-month-olds did not. The shift in prediction skills coincides with the age at which infants learn to perform the action themselves. So, the researchers say, the results are consistent with

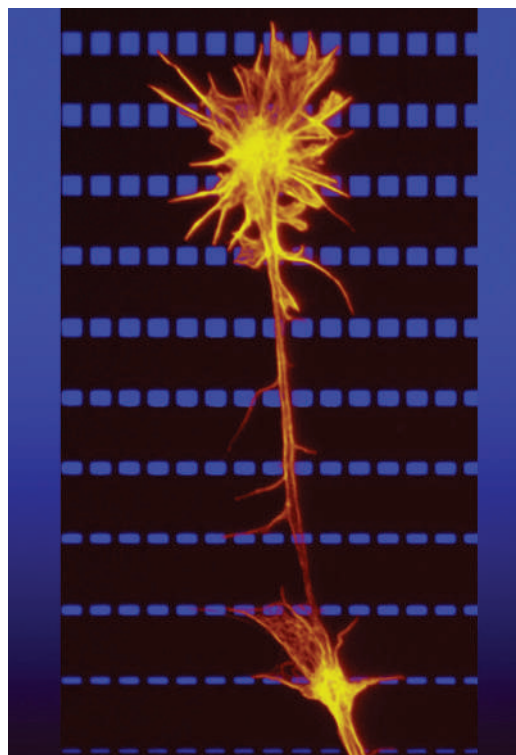


Follow the guidelines

Development **133**, 2487–2495 (2006)
Developmental biologists want to find out how growing neurons wire up correctly using cues from a chemical gradient. Tiny chips printed with guidance molecules may help.

Martin Bastmeyer of the University of Karlsruhe, Friedrich Bonhoeffer of the Max-Planck Institute for Developmental Biology, Tübingen, and their colleagues used a process known as micro-contact printing to effectively create tiny, stripy rubber stamps. They used these to stamp an 'ink' made of ephrinA5 guidance molecules on to glass coverslips (shown blue in picture).

By varying the width and spacing of the stripes, the team shows that chick retinal neurons (yellow) figure out where to stop based on both the local ephrin concentration and the total amount of ephrin encountered on their journey.



A. C. VON PHILIPSBORN

the idea that mirror neurons, trained by use, form the basis of developmental milestones such as understanding others' actions.

PHYSICS

Electron rebellion

Phys. Rev. Lett. **96**, 236804 (2006)

Tantalizing evidence for negative electrical resistance has been gathered by physicists in the United States. The phenomenon, in which an electrical current moves against the applied voltage, is theoretically possible but is hard to observe.

Michael Zudov of the University of Minnesota in Minneapolis and his colleagues looked for the effect using two beams of microwaves to vary the resistance felt by electrons in a semiconductor. The current changed smoothly until one of the beams was tuned to a frequency believed to induce negative resistance, when the current in the semiconductor dropped dramatically. The researchers suggest that this may be evidence of a negative-resistance state interacting with the state set up by the second beam.

NEUROBIOLOGY

Cut to the point

Cell **125**, 1179–1191 (2006)

Exactly how the mutated huntingtin protein causes the hereditary brain condition Huntington's disease has long been a mystery. Now, researchers pin the blame on a process in which an enzyme cuts the protein at a particular point.

Michael Hayden at the University of British Columbia, Canada, and his colleagues studied mice carrying the abnormal huntingtin gene. Previous work had shown that enzymes known as caspases cut the huntingtin protein and linked this to the disease. This team found that, if they altered the gene so that its protein could no longer be cut by caspase-6, the mice did not develop Huntington's. This suggests that the cutting triggers the neurodegeneration typical of the disease and makes caspase-6 a good target for therapy.

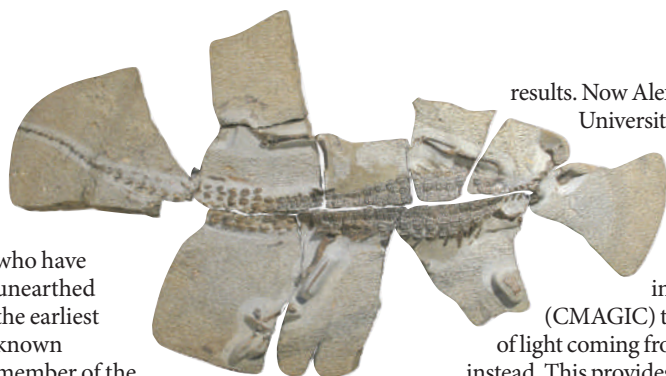
PALAEONTOLOGY

First of the modern crocs

Proc. R. Soc. Lond. B doi:10.1098/rspb.2006.3613 (2006)

The common ancestor of crocodiles and alligators probably arose in the ancient supercontinent of Gondwana, say researchers

S. SALISBURY



who have unearthed the earliest known member of the group in the Australian outback. The fossil (pictured above) dates from between 95 million and 98 million years ago, around the time the group that contains today's species was beginning to diverge.

The new species, *Isisfordia duncani*, was relatively small at just 1.1 metres long, report Steven Salisbury of the University of Queensland in Brisbane and his colleagues. The build of its limbs suggests that the beast was equally at home in and out of water. This flexibility may have helped crocodyliforms to survive the mass extinction that wiped out their fellow reptiles, the dinosaurs, 65 million years ago.

ASTRONOMY

Swell colour

Astrophys. J. **644**, 1–20 (2006)

Two uncertainties that dogged the original discovery of 'dark energy', the unseen force driving the expansion of the Universe, have been deflated by a new analysis.

Measurements of the brightness of Type 1a supernovae, the 'standard candles' of the cosmos, provided the first evidence of the Universe's ballooning. But astronomers were unsure whether more mundane phenomena, such as intergalactic dust or a fundamental difference in character between supernovae now and those in the past, might explain the

results. Now Alex Conley of the University of Toronto, Canada, and his colleagues from the Supernova Cosmology Project use a technique known as colour-magnitude intercept calibration (CMAGIC) to analyse the colour of light coming from such supernovae, instead. This provides independent confirmation that the Universe is swelling in line with current dark-energy predictions.

ELECTROCHEMISTRY

Post production

Nature Mater. doi:10.1038/nmat1672 (2006)

A rusty electrode doesn't sound terribly good news for a battery, but Patrice Simon of the Paul Sabatier University in Toulouse, France, and his co-workers find that an iron oxide can work wonders for rechargeable lithium cells — if the material is nanostructured.

With a negative electrode made from a forest of copper nanopillars coated with Fe_3O_4 , their lithium battery produces six times more power per unit area than one in which the electrode is a flat slab of the same material — which in turn beats the capacity of the graphite electrodes in current commercial cells.

A porous ceramic membrane provides a template for the copper posts, and the product is cheap and non-toxic.

PALAEONTOLOGY

Ancient embryos

Science **312**, 1644–1646 (2006)

Bilaterally symmetric animals existed much earlier than once thought, say researchers. Fossilized embryos indicate the presence

of developmental processes unique to such animals some 40 million years before the Cambrian period.

Jun-Yuan Cheng of Nanjing University and his colleagues isolated 248 embryos from the Precambrian Doushantuo Formation deposits in Guizhou Province, China. The 580-million-year-old embryos were preserved performing their first cell divisions, which feature a lobed structure protruding from one of the cells as it divides. Some modern creatures, including certain molluscs, still use this strategy.

The lobe delivers molecules to one of the resulting cells of an early division, which turn it into a different kind of cell from its counterpart.

GENOMICS

World's smallest genome

Proc. Natl Acad. Sci. USA **103**, 9566–9571 (2006)

With just 373,000 base pairs and a mere 331 genes, it's the smallest genome known — and it has just been sequenced.

Researchers led by Geoffrey McFadden of the University of Melbourne, Australia, have unravelled the DNA of the 'nucleomorph' of *Bigeloviella natans*, a single-celled organism of the group known as chlorarachniophytes.

The nucleomorph is an evolutionary vestige that was originally the nucleus of a eukaryotic cell. The eukaryotic cell swallowed a cyanobacterium to acquire a photosynthetic 'plastid' organelle, and that cell was in turn engulfed by another cell to produce *B. natans* as we know it. Now, most of the nucleomorph's genome is concerned with its own maintenance, and just 17 of its genes still exert any control over the plastid. Its small size suggests it is heading for evolutionary oblivion.

JOURNAL CLUB

Charles Sawyers
Howard Hughes Medical
Institute, University of
California, Los Angeles

One hit or two? A cancer biologist reviews the shifting focus in drug discovery.

Spurred by the success of imatinib — the anticancer drug better known as Gleevec — most drug-discovery companies now have programmes to look for similar,

molecularly targeted therapies.

Imatinib blocks a certain class of enzyme, the tyrosine kinases. In the search for other kinase inhibitors, I am struck by the shift from a laser-like focus on the selectivity of the molecule's binding to interest in molecules that have several targets.

The rationale comes, in part, from the recent clinical success of drugs that target multiple kinases. But is this approach justified?

My views are heavily influenced by clinical studies showing that resistance to kinase inhibitors

occurs principally through mutations that impair drug binding. The mutations are always tracked to a single kinase. For example, resistance to imatinib in leukaemia patients arises exclusively through mutations in the kinase ABL, despite the fact that the kinases KIT and PDGFR are also inhibited. I see this as genetic evidence that the therapeutic effect can be explained by the drug's action on a single target.

However, a recent report (Qi-Wen Fan *et al. Cancer Cell* **9**, 341–349; 2006) convinces me that

hitting two targets may sometimes be better than hitting one.

In a study of various PI(3)-kinase inhibitors, the authors noted that antitumour efficacy requires inhibition of both PI(3)K α and another kinase, mTOR. The logic underlying this result is particularly compelling because both kinases function in the same signalling pathway with multiple levels of feedback.

This study gives me hope that we can move from accidental to rational multi-kinase-inhibitor drug design.

NEWS

Tissue-sample payments anger lawmakers

Revelations about the unauthorized transfer of tissue samples by a US neuroscientist are adding to the challenges faced by researchers who are attempting to set up biobanks — stores of human tissues that can be used to probe the causes of disease.

The latest problems stem from the case of Trey Sunderland, chief of the Geriatric Psychiatry Branch of the National Institute for Mental Health. He is accused of shipping more than 3,000 samples of human spinal fluid to the drug company Pfizer without approval from his employers.

A subcommittee of the US House Committee on Energy and Commerce, which is investigating the shipments, said on 13 June that Sunderland received \$285,000 for the samples. This was part of more than \$600,000 he received from the company between 1998 and 2004.

Sunderland invoked his right not to testify against himself at a 14 June hearing. His lawyer did not return phone calls seeking comment.

The allegations, which surfaced after a tip-off from a whistleblower in April last year, angered lawmakers. "What we have learned from this investigation is that the National Institutes of Health (NIH) lacks adequate controls for human tissue samples, human subject protection, and the scientific conduct of many of its senior employees," says committee member John Dingell (Democrat, Michigan).

Such criticism worries researchers, because

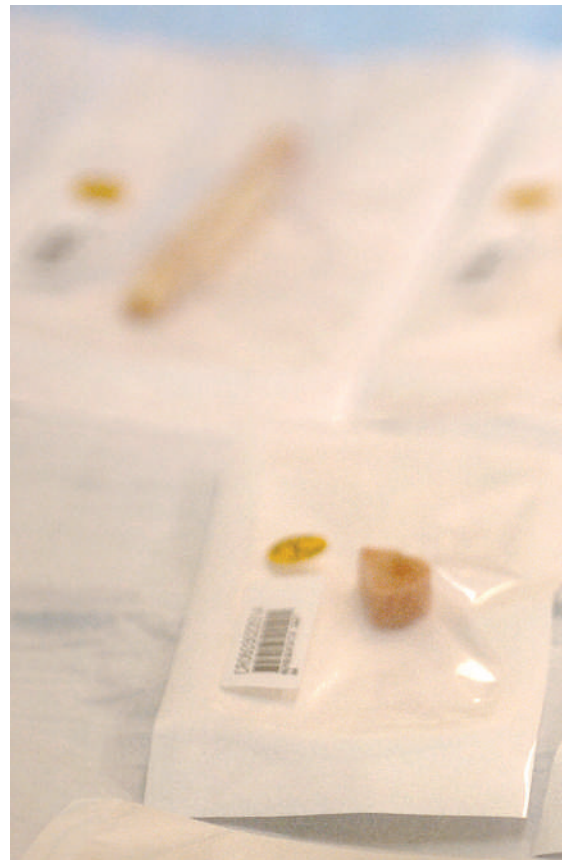
scientists are starting to use stored banks of tissue to link genetic, medical and personal health data. Iceland, Estonia and the United Kingdom are already developing national databases, and the United States is considering setting up a biobank containing information on hundreds of thousands of Americans (see *Nature* **429**, 475–477; 2004). The NIH has already announced two specific projects linking genetic and environmental information on patients with conditions such as Alzheimer's disease.

Ethicists and scientists fear that reports of lax policies and misconduct will hurt such initiatives. "If there is a perception that the researcher is using the samples to line his or her pockets on the side, you're going to lose public support,"

says Arthur Caplan, an ethicist at the University of Pennsylvania in Philadelphia. "There is nothing more corrosive to altruism than people making money off research."

Scandals elsewhere have already had similar effects. In Britain, for example, scientists reported a drop in the donation of organs for research after a pathologist at Alder Hey Hospital in Liverpool was revealed to have illegally removed organs from the bodies of children on which he had conducted post-mortems (see *Nature* **409**, 655; 2001). The incident led to new human-tissue legislation in 2004 and the creation of the Human Tissue Authority, tasked with overseeing the handling of such samples.

"There is nothing more corrosive to altruism than people making money off research."



Plans to set up human tissue banks have been hit by the alleged sale of samples to a US drug firm.

In Sunderland's case, a lack of oversight is giving particular cause for concern. "A lot of this stuff was collected with pretty vague consent and rules," says Caplan. "It's an issue and it's going to pop up again." Michael Gottesman, NIH deputy director for intramural research, says the agency will clarify its policies on the storage and shipment of tissue samples and develop an electronic database for tracking

Doomsday food store takes pole position

There's something fitting about the decision to site a bastion against the end of the world in a place that looks as if it has already experienced the apocalypse. On 19 June, a construction crew started work on a doomsday seed bank from which the genetic riches of Earth's food crops could, if necessary, be reconstituted. The location: the island of Spitsbergen in the Svalbard archipelago, a desolate place where

the winters are long and dark, and polar bears outnumber people.

The island's remoteness, say staff at the Global Crop Diversity Trust, a charitable organization based in Rome that has helped develop the bank, makes it the ideal location to store samples safely. Spitsbergen is free of tectonic activity, and its permafrost would preserve seeds at around -4°C . Coal from a local mine will be used to power

refrigeration units that will further cool the bank to the internationally recommended standard of -20 to -30°C .

"It's a resource that needs to last for ever," says Cary Fowler, the trust's director. Fowler says that such a bank is needed in case a species becomes extinct or loses genetic variety through overuse. Some of the three million seeds earmarked for storage could be

called upon to thwart a new disease or drought, or even extinctions caused by nuclear war. Although other seed banks exist, the trust is developing the Spitsbergen facility because many banks are poorly maintained and face uncertain futures owing to funding problems.

Norway, which administers Svalbard, will fund the US\$3-million construction costs, and the trust will pay for its maintenance, which



AP PHOTO/N. BERGER

patient samples in all of its 27 branches.

Such measures could help to restore confidence among potential donors. But biobank officials will have to tackle other issues if the initiatives are to succeed, such as questions about who 'owns' the information stored in tissue repositories. In April, for example, a US federal judge had to intervene in a dispute between patients, a researcher and the University of Washington in St Louis over a valuable tissue repository housed at the university (see *Nature* **440**, 1102–1103; 2006).

There are also concerns about how investigators protect the privacy of patients who donate tissue, and how they give consent for future studies on these tissues. National laws differ on such matters, although several efforts are under way to devise uniform rules. "There's a real movement for harmonization across international barriers," says Mark Sobel, executive officer of the American Society for Investigative Pathology. "I think by 2007 or 2008, we're really going to see some global acceptance for it." ■

Erika Check

Fowler estimates at around \$100,000 annually. The trust, which expects the facility to be completed by 2007, will also help developing countries prepare and transport seeds to the Arctic.

The bank will be carved into one of the island's sandstone mountains, and will consist of a 50-metre tunnel leading to a storage facility reinforced with

one-metre-thick concrete. Seeds will be wrapped in aluminium foil to keep out moisture. The cave will be protected by two high-security doors armed with motion detectors. No full-time staff will oversee the facility, says Fowler, because it is accessible only by an air-strip about three kilometres away, making it relatively easy to track people's comings and goings.

"It's about providing long-term security for crop plants," says plant scientist Matthew Daws of the Millennium Seed Bank Project at Kew Gardens in London. "It's an insurance policy for countries to deposit some of their collection and have confidence that after thousands of years their seeds will be viable." ■
Jacqueline Ruttimann

ON THE RECORD

"Life on Earth is at the ever-increasing risk of being wiped out by a disaster such as sudden global warming, nuclear war, a genetically engineered virus or other dangers we have not yet thought of."

Cosmologist Stephen Hawking delivers an upbeat assessment of why humans should colonize space.

Source: AP

SCORECARD



Dirty rats

A study of sewer rats reveals that they have healthier immune systems than their hygienically protected laboratory cousins, leading scientists to ponder the medical value of filth.



David Beckham

A branding study shows that British children are obsessed with celebrities. Top of the pile is the England football captain, who is more on the kids' minds than the toys and clothes marketed at them.



Sunscreen

US toxicologists find that nanoparticles used in some sunscreens and cosmetics might be able to cause damage to nerve cells, at least in mice.

OVERHYPED

Chinese engineers

Last autumn, the US National Academies set off alarm bells in Washington with a report claiming, among other things, that China had produced 600,000 engineering graduates last year to America's 70,000. The numbers were compelling enough to help convince President Bush to endorse a multi-billion-dollar 'competitiveness initiative'.

But this week, the academies quietly revised the China number to 350,000 and the US number to 140,000. Why? It seems that the original report was comparing apples with oranges. Or in this case, fully fledged US engineers with the Chinese equivalent of car mechanics.

Open-access journal hits rocky times

The Public Library of Science (PLOS), the flagship publisher for the open-access publishing movement, faces a looming financial crisis. An analysis of the company's accounts, obtained by *Nature*, shows that the company falls far short of its stated goal of quickly breaking even. In an attempt to redress its finances, PLOS will next month hike the charge for publishing in its journals from US\$1,500 per article to as much as \$2,500.

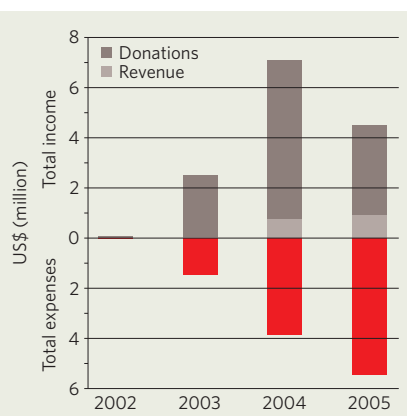
Subscription-based journals recover much of the costs of peer review and editing — and in the case of commercial publishers such as Nature Publishing Group, make some of their profits — by charging for access to their products. But the PLOS journals, the first of which launched in 2003, adhere instead to an 'author pays' open-access model: costs are recovered by charging authors, and papers are made available free of charge to the end-user.

The publisher's premiere titles, *PLoS Biology* and *PLoS Medicine*, are intended to compete on quality with high-ranking weekly subscription journals such as *Nature*, *Science* and *The New England Journal of Medicine*. An equal goal was to show that the economics of 'author pays' could be made to work, by making PLOS financially sustainable.

But although *PLoS Biology* has achieved an impact factor of 14.7, a more than respectable score for a relatively new journal, an analysis of PLOS's accounts shows that the financial side of the business looks less rosy. As a US non-profit charity, PLOS must file its annual accounts to the Internal Revenue Service. *Nature* consulted these via GuideStar.org, a database that contains information on 1.5 million US non-profit organizations.

The figures show that PLOS lost almost \$1 million last year. Moreover, its total income from fees and advertising currently covers just 35% of its total costs. And although this income is increasing — from \$0.75 million in 2003–04 to \$0.9 million in 2004–05 — it lags far behind spending, which has soared from \$1.5 million to around \$5.5 million over the past three years.

To stay afloat, the firm continues to rely on the philanthropic grants that launched the project: \$9 million from the Gordon and Betty Moore Foundation and \$4 million from the Sandler Family Supporting Foundation, both based in San Francisco (see graphic). These covered 65% of the company's operating costs



The ups and downs of open access: PLOS faces a financial balancing act.

last year, but are running out: at the end of last September, PLOS had assets of \$3,393,265.

"We will continue to rely on philanthropic grant support for the foreseeable future," says Mark Patterson, director of publishing at PLOS's UK office in Cambridge, and "possibly always". Patterson adds that he is hopeful that the Sandler Foundation will provide more grants.

"This demonstrates once again the fragility of the author-pays model," says David Worlock, chairman of the London-based publishing consultancy Electronic Publishing Services. (Worlock has worked with a number of publishing companies including Nature Publishing Group.) "It's a real giveaway if they are now saying that they will always need some philanthropic funding."

But Patterson points out that PLOS launched most of its journals recently, and that income from these publications is only beginning to accrue. "The financial situation for this year will look quite different," he says. "I'm confident we can balance the books this year and next."

The three PLOS journals that launched last year — *PLoS Computational Biology*, *PLoS Genetics* and *PLoS Pathogens* — are run by vol-

unteer academic editorial teams, rather than in-house staff. This should help to ease the pressure on finances overall, explains Patterson, as they have lower editorial costs and are more likely to break even than *PLoS Biology* and *PLoS Medicine*.

In a further move towards high-volume publishing, PLOS intends to launch an open-access online database, called PLOS One, later this year. Papers published on the database will be peer reviewed only to check that they are technically sound, and not to assess novelty.

The resource will publish "hundreds of papers per month," says Chris Surridge, its Cambridge-based managing editor, and pay for itself through fees and economies of scale.

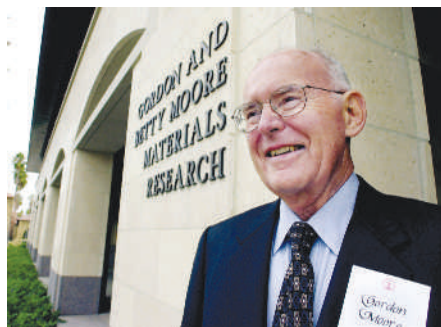
The diversification should help PLOS balance its books, agrees Matthew Cockerill, publisher of BioMed Central, a London-based open-access publisher. The high cost and low volume of articles in PLOS's flagship journals mean "they are not by themselves financially sustainable in isolation," he argues.

BioMed Central, which last July increased its charges from \$500 to as much as \$1,700, depending on the journal, has also yet to break even. It is "getting close," says Cockerill, who declined to give an estimate of when break-even would occur: "Predicting the future is not a wise activity to engage in."

The hike in PLOS's fees, announced earlier this month, is due to take effect from 1 July. It is similarly intended to improve the firm's financial position by "reflecting more closely the costs of publication," according to a statement on the PLOS website announcing the increases. Author charges will rise to \$2,500 for *PLoS Biology*, *PLoS Medicine* and *PLoS Clinical Trials* and to \$2,000 for *PLoS Computational Biology*, *PLoS Genetics* and *PLoS Pathogens*.

When PLOS launched, other publishers expressed scepticism over its sustainability, arguing that the \$1,500 fee was unrealistically low for covering the costs of producing a high-quality journal (see *Nature* 425, 554–555; 2003). PLOS rejected this argument as coming from journals with a vested interest in maintaining the reader-pays status quo. It said the \$1,500 figure was realistic and was based on a business plan thrashed out with publishing experts. "Four years further on, we know a lot more," says Patterson.

Declan Butler



Gordon Moore's philanthropic foundation has donated \$9 million to the PLOS publishing effort.

P. SAKUMA/AP


**EUROPEAN SOCIETY OF
HUMAN REPRODUCTION
AND EMBRYOLOGY**

 Discover the latest finds in
reproductive research.

www.nature.com/news

Congress pushes plan to make papers free

A question to holders of grants from the National Institutes of Health (NIH): do you plan to submit publications resulting from NIH-funded research to PubMed Central? The answer is probably 'no'. Fewer than 4% of NIH-funded researchers send their papers to the free-to-access archive, despite the agency having requested that they do so since May 2005.

That could be about to change. The archive is at the heart of plans, backed by a coalition of US politicians and advocacy groups, to make the fruits of publicly funded research freely available. On 13 June, a change was introduced to a House of Representatives spending bill such that NIH grantees would be required, not requested, to submit to PubMed within 12 months of publication. The proposal has upset some publishers and scientific societies, who are wary of citation confusion and a possible drop in income.

The move is in part a response to the limited impact of NIH's current policy on open access. "It is not working in its current state," says Norka

Ruiz Bravo, NIH deputy director for extramural research and an advocate of full participation in open-access publishing by all grantees.

A majority of the NIH Public Access Working Group, an advisory body that includes publishers and researchers, have also endorsed mandatory submission to PubMed Central, says Thomas Detre, a psychiatrist at the University of Pittsburgh in Pennsylvania and chair of the working group: "Nobody in academia will do anything administrative that's not obligatory."

But critics charge that such a policy could affect the integrity of the scientific literature. Some journals do not allow copyedited versions to be posted on PubMed Central. This means two versions of the same research paper can be published: a peer-reviewed manuscript version held in PubMed Central and a journal version complete with copyedits and other mostly cosmetic modifications. The PubMed Central ver-

sion would also not necessarily link to the journal that published the paper. This would unfairly divert web traffic from the journal, and the multiple versions could confuse citation analyses, says Martin Frank, executive director of the

American Physiological Society in Bethesda, Maryland.

Such arguments could be addressed when the Senate drafts its appropriations bill,

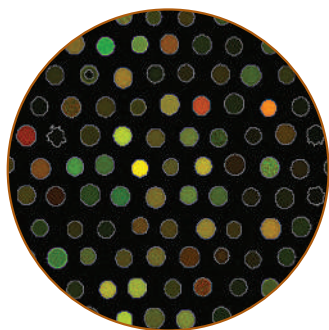
as some senators are likely to want to include a clause on mandatory submission. In May, for example, Joseph Lieberman and John Cornyn introduced a bill that would require that all federal US agency grantees with annual research budgets of more than \$100 million make their research papers freely available within six months of publication (*Nature* 441, 140; 2006).

The Senate's spending bill could be introduced in July, although completion by the autumn is more likely.

Gene Russo

**"Nobody in academia will
do anything administrative
that's not obligatory."**

More Published



In a recent PubMed survey, more publications referenced GenePix® from Molecular Devices than any other slide-based microarray scanner platform. Why? Because researchers trust the results they get with GenePix scanners and Acuity® analysis software. When you trust the tools you use, you have the confidence to publish your results.

Obtaining reliable data requires high performance instruments and software that are easy to use and increase your productivity. With Molecular Devices' microarray scanning solutions, we give you that and more—so you can get more published.

Microarray Scanning

- ☉ GenePix Autoloader 4200AL: 36-slide capacity with automation
- ☉ GenePix Professional 4200A: four-laser flexibility and upgradability
- ☉ GenePix 4000B: fastest two-laser scanner in existence
- ☉ GenePix Personal 4100A: affordable, quality two-laser scanning

Microarray Analysis Software

- ☉ GenePix Pro 6 microarray image analysis software
- ☉ Acuity 4 microarray informatics software

 **Molecular Devices**

tel. +1-800-635-5577 | www.moleculardevices.com

Expect more. We'll do our very best to exceed your expectations.



**NANOPARTICLES
STRESS BRAIN CELLS**
Tiny grains send cells
into potentially
dangerous overdrive.
www.nature.com/news

Plan to rank universities fails to impress

European nations have long cast envious glances across the Atlantic at the industrial spin-offs — and the associated wealth — created by US universities. But how hard should the continent's governments push when it comes to persuading their universities to become more focused on application? A report released last week shows that Britain is pushing pretty hard, and critics say the country risks undermining fundamental research in order to subsidize industry-funded projects.

Scientists might have been expected to welcome the proposed changes, which involve reform of the much-criticized Research Assessment Exercise (RAE). The roughly five-yearly process, in which peer review is used to assess and reward individual university departments, is widely acknowledged to have boosted the quality of British science since it was introduced in 1992. But given that the assessment has paid off, and the preparing and processing of the submissions is so time-consuming, the RAE is felt to have had its day.

Yet proposals to make the 2008 RAE the last of its kind are attracting as much criticism as the current mechanism does. In a consultation document released on 13 June, the government suggested that money be allocated to universities based on how much income they already attract from competitive grants, charities and industrial contracts. The system was chosen because, across the whole university sector, this income correlates well with the results of RAE peer review.

But the correlation doesn't always hold at the level of individual universities. Five different models are proposed, all of which have roughly the same effect of boosting universities that are good at bringing in money, such as those with the research capacity to attract significant funding from medical charities (for example, University College London) or industry (such as Cranfield University), and taking money from those for whom such activity is a smaller component of their total income (see "Who would the scheme most affect?").

Poorly rated

Other countries have made use of metrics, but the British initiative is the boldest and most application-focused. Germany, for example, uses income measures to allocate around a quarter of local-government funding for universities, but a range of other metrics, including the number of graduates and PhDs, are also considered.

If the British system goes ahead, it will mean



The University of Oxford stands to gain from Britain's plan to reward universities good at winning grants and contracts.

N. HAMID

that "universities are there to do research for other people", says Tom Sastry, a Bristol-based senior researcher at the Higher Education Policy Institute, an independent think-tank. He and other critics fear that this is the intention of the Treasury, which wants to push universities towards application-driven projects; the Treasury first outlined the metrics plans in the budget this March. Fundamental research, which does not attract so much external funding, could then suffer. "Many theoretical fields are going to find it hard," agrees Anthony Van Raan, an expert in research assessment methods at Leiden University in the Netherlands.

Van Raan and others say that the Treasury is right to focus on metrics, but only if a broad range of measures is used. These could include numbers of PhD students trained, together with some form of bibliometric analysis and other output measures, such as patents gained. "The scientific community will be disappointed if the final models are not more sophisticated," says Jonathan Adam, director of

Evidence, a Leeds-based consultancy that analyses the quality of international research.

The Treasury seems, however, to have limited interest in expanding the range of metrics used. Neither of the authors behind the consultation document — David Eastwood, vice-chancellor of the University of East Anglia in Norwich, and Alan Wilson, director-general for higher education in the government's Department for Education and Skills — would comment. But a Treasury spokesman, while noting that the consultation asked for views on the possibility of including other metrics, said that "bibliometrics are not sufficiently developed to have been incorporated into the modelling" and described PhD numbers as a "less robust measure".

Such points are disputed by policy experts. Crude bibliometrics measures, such as the number of citations earned by a department, are accepted as being inadequate because some fields, such as mathematics, tend to use fewer references than others, such as genetics. But Van Raan and Adam have independently developed more sophisticated impact measures that place citation numbers in context within disciplines and between countries, and both say that such metrics should form a part of any university assessment system. Unless the range is expanded, says Van Raan, universities could become overly focused on short-term pay-offs: "This might stop research that could be interesting in ten years or so."

The proposals are open for comment until 13 October. A test drive of the agreed scheme will run alongside the existing RAE scheme in 2008, before being phased in as its replacement.

Jim Giles

WHO WOULD THE SCHEME MOST AFFECT?		
WINNERS		
Average annual funding increase (percentage change)		
1	University of Oxford	£4,783,130 (8)
2	Cranfield University	£3,909,877 (109)
3	Queen Mary, University of London	£3,509,060 (33)
4	University of Greenwich	£3,484,424 (255)
5	University College London	£2,293,247 (4)
LOSERS		
Average annual funding decrease (percentage change)		
1	University of Manchester	£7,110,984 (15)
2	Imperial College London	£4,437,005 (9)
3	University of Cambridge	£3,968,663 (6)
4	University of Southampton	£3,577,707 (13)
5	University of Sheffield	£3,360,936 (12)

SPECIAL REPORT

Lure of lie detectors spooks ethicists

US companies are planning to profit from lie-detection technology that uses brain scans, but the move to commercialize a little-tested method is ringing ethical and scientific alarm bells. **Helen Pearson** reports.

Bioethicists and civil-rights activists are calling into question plans by two US companies to single out liars by sliding them into a brain scanner and searching their brains for give-away patterns of deception.

The two firms say that they will give the accused a chance to prove their innocence using a technique more accurate than the discredited polygraph. No Lie MRI will start offering services out of Philadelphia this summer. Those behind the second company, Cephos, based in Pepperell, Massachusetts, say they hope to launch their technology later this year. Likely clients include people facing criminal proceedings and US federal government agencies, some of which already use polygraphs for security screening.

Critics say that the science underlying the companies' technique is shaky and that the premature commercialization of the method raises ethical concerns about its eventual use in interrogation. This week, the American Civil Liberties Union (ACLU) entered the debate by organizing a 20 June briefing on the issues for scientists, the public, the press and policy-makers in Washington DC.

The field of lie detection is littered with dubious devices. The polygraph relies on the idea that lying is stressful, and so measures changes in heart rate, breathing and blood pressure. But because it can be duped by countermeasures and there is little hard evidence that it actually works, it is rarely admitted as evidence in court.

Rather than relying on indirect measures of anxiety, assessing brain activity using functional magnetic resonance imaging (fMRI) goes to the very source of the lie. In one of the earliest studies, a team led by Daniel Langleben of the University of Pennsylvania, Philadelphia, and his colleagues offered students sealed envelopes containing a playing card and \$20. The students were told they could keep the money if they could conceal

which card they held when questioned in an MRI machine (D. D. Langleben *et al. Neuro-Image* 15, 727–732; 2002).

These and other studies revealed that particular spots in the brain's prefrontal cortex become more active when a person is lying. Some of these areas are thought to be involved in detecting errors and inhibiting responses, backing the idea that suppressing the truth involves additional areas of the brain to telling it.

The early studies showed that it was possible to make out subtle changes in brain activity caused by deception using pooled data from a group of subjects. But to make a useful lie detector, researchers must be able to tell whether an individual is lying; when only one person is assessed it is much harder to tease out a signal from background noise. Langleben, who advises No Lie MRI, says he is now able to tell with 88% certainty whether individuals are lying (see *Nature* 437, 457; 2005). A group working with Cephos, led by Andrew Kozel, now at the University of Texas Southwestern Medical Center in Dallas, makes a similar claim.

Thought police

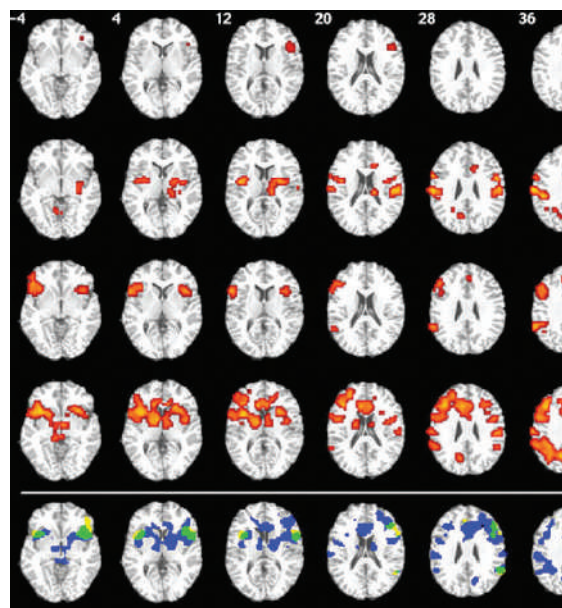
Kozel and his colleagues asked 30 subjects to take

either a watch or a ring, hide it in a locker and then fib about what they had hidden when they were questioned inside a scanner. Using the results of this study, the team devised a computer model that focuses on three regions of the brain and calculates whether the shift in brain activity indicates lying. When the model was tested on a second batch of 31 people, the team reported that it could pick up lies in 90% of cases (F. A. Kozel *et al. Biol. Psychiatry* 58, 605–613; 2005).

But critics of the technology urge restraint. "Until we sort out the scientific, technological and ethical issues, we need to proceed with extreme caution," says Judy Illes of the Stanford Center for Biomedical Ethics, California.

"Until we sort out the scientific, technological and ethical issues, we need to proceed with extreme caution."

TEK IMAGE/SPL



One problem is that there is no standard way to define what deception is or how to test it. Scientists also say that some of the statistical analyses used in the fMRI studies are questionable or give results that are perilously close to the thresholds of significance. "On individual scans it's really very difficult to judge who's lying and who's telling the truth," says Sean Spence of the University of Sheffield, UK, who was one of the

LANGLEBEN ET AL.

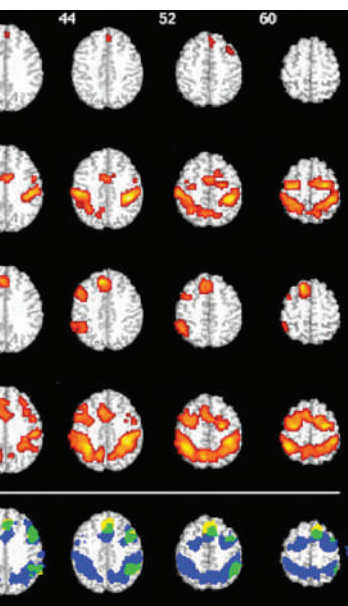


MAKING THE MOST OF A LITTLE DNA

New tests could make it easier to detect IVF embryos at risk of genetic disease.

www.nature.com/news

HYBRID MEDICAL ANIMATION/SPL



Caught out: lie detectors that use MRI scans (left) capture brains in the act of deception, whereas conventional methods rely on stress-induced changes such as fluctuating pulse rates (above).

be at stake. They say there are no data about whether the technique could be beaten by countermeasures, and that data collected from healthy subjects reveal little about the mindset of someone who genuinely believes they are telling the truth or someone who is confused, delusional or a pathological liar.

"If I'm a jihadist who thinks that Americans are infidels I'll have a whole different state of mind," says Gregg Bloche, an expert in biomedical ethics and law at Georgetown University Law Center, Washington DC, and a member of the ACLU panel. "We don't know how those guys' brains are firing."

Because of these concerns, legal experts say that the technology is unlikely to pass the standards of scientific accuracy and acceptance required for it to be admissible in a US court. But even if it is not sufficiently accurate and reliable today, it may well be tomorrow, as more and more people are tested and techniques refined. That raises a second set of concerns that revolve around who should be allowed to use the technique and under what circumstances.

Bioethicists worry that fMRI lie detection could quickly pass from absolving the innocent to extracting information from the guilty — in police questioning, immigration control, insurance claims, employment screening and family disputes. Their concerns are fuelled by other emerging lie-detection technologies, such as those that measure the brain's electrical activity (see *Nature* 428, 692–694; 2004).

Truth be told

Particularly in the aftermath of 11 September 2001, they worry that fMRI and other devices might be misused in the hands of the military or intelligence agencies. "There's enormous pressure coming from the government for this," says bioethicist Paul Root Wolpe at the University of Pennsylvania. "There is reason to believe a lot of money and effort is going into creating these technologies."

On top of this, ethicists say there is something deeply intrusive about peering into someone's brain in search of the truth; some even liken it to mind-reading. In future, they say, a suspect might be betrayed by their prefrontal cortex before they even open their mouth — if, for example, the brain recognizes a particular photo or foreign word. "This is the first time that we have ever been able to get information directly from the brain. People find the idea extraordinarily frightening," Wolpe says.

No Lie MRI founder Joel Huizenga and

Cephos head Steven Laken say they are aware of both the scientific limitations and the ethical concerns. Laken says that the company only plans to use the technique with people and questioning protocols that are as similar as possible to those used in the study, and that he wants to work with attorneys to iron out problems along the way. "We really want to get the science right," Laken says. "We don't want to get to court and be killed."

In terms of the ethics, they point out that the test can only be used on those who consent — because an unwilling subject could easily foil the fMRI machine by simply moving or refusing to answer the questions. (Critics counter that just declining an fMRI test could be incriminating.) No Lie MRI has a licensing agreement for the technology from the University of Pennsylvania and Langleben says he would "yank their licence" if the company overstepped ethical boundaries.

Whatever the objections, the two companies say they are already receiving numerous enquiries from people eager to prove their innocence. Huizenga says that he has eight TV shows lined up to document some of the first customers' slide into the machines. The ultimate goal, he says, is to have franchises all over the world. These would collect MRI scans and beam data to the company's computers for central analysis. The company will charge \$30 per minute for the scans, which might take one hour, plus additional fees for legal assistance and developing questions.

Some researchers feel that such plans are of only limited cause for concern. "Most people who do brain imaging think this is far too soon

and that this will crash and burn," says Spence. "So it's not worth getting worked up about."

But bioethicists maintain that there needs to be far greater discussion, both within and beyond the scientific community, before the technology is

unleashed. Stanford law professor Hank Greely organized a March workshop on lie detection and the law and is also a member of the ACLU panel. He suggests that an impartial agency should introduce a regulatory scheme that would prevent the use of MRI for lie detection until there was sufficient evidence to conclude that it was proven safe and effective — much as the US Food and Drug Administration bars or approves a drug.

Bioethicists add that neuroscientists in particular need to flag up some of the social and legal issues if they are to avoid fMRI earning a bad label. "The scientists involved in this have an absolute obligation to shepherd this technology," Wolpe says.

first to publish on the use of MRI in the study of deception. "The studies might not stand up to scrutiny over the long term."

Another concern raised by scientists and bioethicists is that the contrived testing protocols used in the laboratory — in which subjects are told to lie — cannot necessarily be extrapolated to a real-life scenario in which imprisonment or even a death sentence could

"This is the first time that we have ever been able to get information directly from the brain."

Satellite aims to detect antimatter in cosmic rays

The first satellite built to detect antimatter in space launched safely last week. The PAMELA probe took off from the Baikonur Cosmodrome in Kazakhstan on 15 June, carrying instruments that will detect antiprotons and positrons, the mirror particles of protons and electrons. The hope is that it could help identify the mysterious 'dark matter' that makes up more than 80% of the Universe.

Cosmic rays formed of high-energy particles constantly bombard Earth. The particles, which include antimatter, are almost entirely absorbed by the atmosphere, and attempts to survey them have been limited to short balloon flights. PAMELA will spend three years in space, and should detect thousands of antiparticles during its mission.

Death penalty can be challenged as 'cruel'

US prisoners on death row can object to the way the state plans to kill them as "cruel and unusual punishment", according to a

unanimous judgement by the US Supreme Court on 12 June.

Such punishments are banned by the constitution. But individual courts might still determine that the current method of lethal injection — which some scientists claim could be painful — is not cruel and unusual.

In essence, the ruling simply clears the way for the courts to examine the scientific evidence. Nevertheless, anti-death-penalty activists see the decision as a win. "As cases

reveal the risk of excruciating pain in lethal injection procedures, states may come to realize that any method of execution is cruel and unusual," says Sarah Tofte, a consultant at Human Rights Watch, based in New York.

DNA researcher found guilty of falsifying data

The US Office of Research Integrity has found DNA-repair researcher Steven

Endangered rhino stumbles into the limelight

This mysterious passerby is one of the very last rhinoceroses on the island of Borneo. It is a subspecies of the critically endangered Sumatran rhino (*Dicerorhinus sumatrensis*). The photo — the first snap of a Borneo rhino — was taken with a motion-triggered camera as part of an effort to document and protect the population, which is believed to number 13. It seems a bleak hope that this population can be saved from the threat posed by poachers and the destruction of its forest



habitat. The conservation agency WWF and Malaysia's Sabah Wildlife Department jointly unveiled the picture on 13 June.

WWF-MALA SIA/R. ALFRED

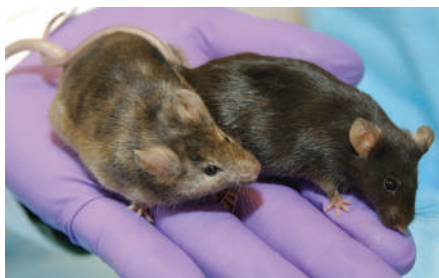
Leadon guilty of scientific misconduct, it announced on 8 June. Leadon, formerly a professor at the University of North Carolina at Chapel Hill, was previously found guilty of falsifying research findings by a university panel in 2003. The US federal finding said that eight publications and several grant applications were affected.

Previous retractions of Leadon's papers left the field in disarray (see *Nature* 435, 1015; 2005). He has now signed an agreement that forbids him from receiving federal grants or consulting for the government for five years. He must also ask to retract three more articles.

Leadon denies any misconduct. "I did not engage in scientific misconduct — an error was discovered in the experimental protocol," he wrote in a statement e-mailed to *Nature*. He says he entered into a settlement because he cannot afford the legal costs of fighting his case.

NIH cash boosts bid for more knockout mice

Two public mouse repositories last week received \$800,000 from the National Center for Research Resources at the US National Institutes of Health (NIH) in Bethesda,



It's a knockout: NIH funds will make another 300 types of mouse mutant available to researchers.

Maryland. The move is part of a plan to make genetically modified 'knockout' mice more easily accessible to researchers.

Knockout mice have become a powerful tool in basic and medical research. But three-quarters of the approximately 4,000 knockout mouse lines described in the scientific literature have not yet been placed in public repositories. The money will be distributed to the Mutant Mouse Regional Resource Centers at the University of California, Davis, and the Harlan facility of the University of Missouri, Columbia, with the goal of making available more than 300 existing mouse mutants. These repositories will help an NIH project that aims to create a knockout mouse for each of the 20,000 genes in the mouse genome.

Japan steps up attempt to overturn whaling ban

Japan has won an unprecedented, if largely symbolic, victory at the International Whaling Commission (IWC).

On 18 June, members passed a resolution noting "concern that the IWC has failed to meet its obligations". Japan feels that a 1986 moratorium on commercial whaling has continued "irrespective of stock conditions" and in opposition to the original goal of the commission, says Hideki Moronuki, a spokesman at Japan's fisheries ministry.

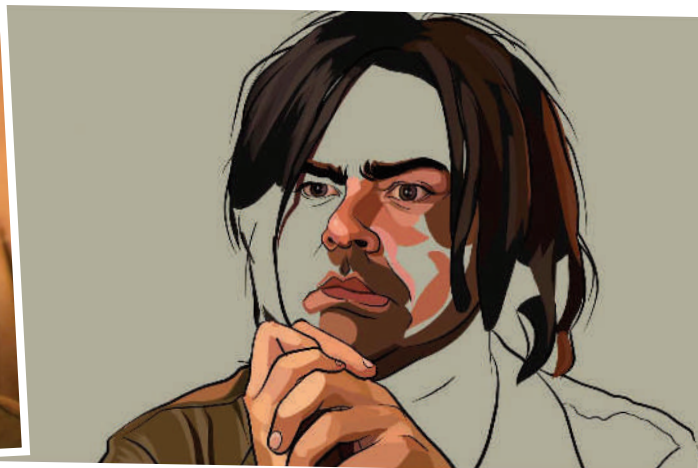
The resolution calls for a "normalization" of the commission's operations, and notes that "the moratorium which was clearly intended as a temporary measure is no longer necessary". But support for the resolution, a narrow 33–32 vote, falls far short of the three-quarters majority needed to overturn the moratorium. "It's a first step," says Moronuki.

Correction

The News story 'Born or made? Debate on mouse eggs reignites' (*Nature* **441**, 795; 2006) should have said that the mice in Jonathan Tilly's study expressed GFP in egg cells whereas the mice in Amy Wagers' study expressed GFP in all cells.

FROM MICROSCOPE TO MULTIPLEX

There's more to science at the movies than Lex Luthor's attempts to synthesize kryptonite. In the first of two features on film, **John Whitfield** looks at how a cinematographic technique can provide insights into the perception of reality. In the second, **Alison Abbott** meets Ben Heisenberg, a director whose first film is a taut moral fable of laboratory life.



An MRI scanner darkly

What is reality? And does anybody care? These questions permeate the work of science-fiction writer Philip K. Dick, whose characters' memories and identities are frequently products of drugs and other technologies — implying that recollections, as well as appearances, should never be trusted. Dick's themes have been catnip to film-makers — *Blade Runner*, *Total Recall* and *Minority Report* are all based on his stories. The latest example, *A Scanner Darkly*, directed by Richard Linklater, will be released this July.

Neuroscientists often ask the same questions. And one of the tools that they are using to address them is 'rotoscoping', a filming technique Linklater uses in *Scanner* and some of his earlier work to blur the distinction between reality and animation. Our notions of authenticity, brain scientists are finding, depend as much on emotional and psychological plausibility as they do on physical accuracy — and the brain will swallow almost anything, provided it comes in the form of a story.

"Our ability to think of others as having minds is very promiscuous, and applies itself across a wide range of entities," says neuroscientist Rebecca Saxe of the Massachusetts

Institute of Technology. Watch a crude animation of a big square tracking a small one, for example, and the word 'chasing' springs to mind. Art exploits this promiscuity, creating emotional impact from strings of abstract symbols such as the lines of letters in novels, or from flickering images on screens.

Studies of brain activity using functional magnetic resonance imaging (fMRI) show that the brain distinguishes movements that might seem to have intentions from other movements. Seeing mechanical motion, such as that of a pendulum, involves a different part of the brain from seeing biological motion, even if it's only that of a cartoon arm.

But how do neuroscientists know whether our responses to such lab-based fictions reflect the real world? "I was interested in whether people attributed intentions to cartoon people in the same way that they comprehend the intentions of real people, but finding or creating videos with cartoon and real people doing the exact same thing seemed very difficult," says neuroscientist and keen filmgoer Raymond Mar of the University of Toronto in Canada. Mar struck lucky; he discovered that Linklater's earlier rotoscope movie, *Waking Life* (2001) provided just

such material. In rotoscoping, conventionally shot footage is transformed into cartoonish animation using a combination of human animators and computers; *Waking Life* uses the technique as a way of presenting a constantly shifting world of dreams, whereas *A Scanner Darkly* uses it to evoke the estranged world of drug use. In the extras on the first movie's DVD, Mar found clips of the video footage that had been transformed into animation, ranging from mellow scenes of pillow talk between Julie Delpy and Ethan Hawke to a man tipping petrol over himself.

Fact or fiction?

Mar and his colleagues showed both cartoon and video clips to subjects in an MRI scanner. They found that two areas of the cortex previously associated with detecting intention, the superior temporal sulcus and the temporal parietal junction (TPJ) fired more strongly in response to the video footage than to the cartoons¹. Cartoon sequences, on the other hand, produced more activity in an area called the bilateral orbitofrontal cortex, which responds to rewarding stimuli. Mar's team speculate that this region may have been more tickled by the trippy cartoon footage than the more

© WARNER BROS. PICTURES 2006



Life like: a filming technique in which normal footage is transformed into cartoons is being exploited by neuroscientists to assess how the brain responds to more or less realistic scenes.

mundane video; it's certainly the case that watching the rotoscoped dreamscapes is a peculiarly rich experience.

"It's an elegant but challenging study," says Saxe. "It's hard to imagine a more minimal contrast that more effectively manipulates this dimension. But the findings are mysterious." The mystery is what the patterns of brain activity in the two treatments mean. It may be that the lower activity in the TPJ stimulated by the cartoons reflects the fact that the animations are easier to process into ideas about intention than the 'real life' footage. Mar, however, plumps for an alternative, although not necessarily contradictory explanation — that the more detailed video footage contains more cues of intention, ticking more of the brain's boxes. His next plan is to show subjects both a real, live hand poking into the scanner and a video of the same event, and see how their brains respond.

"It suggests a template-matching mechanism," adds neuroscientist Kevin Pelphrey of Duke University in Durham, North Carolina. "It may be that the more realistic stimuli portray more intentionality, which these brain regions prefer." But more realistic doesn't necessarily just equal more convincing. The best graphics and robots risk toppling into what, in 1970, the Japanese roboticist Masahiro Mori dubbed the uncanny valley, where their almost-but-not-quite realness becomes creepy and repellent.

Earlier studies have indicated that the brain's mind-reading areas work harder if they believe they are perceiving a real person. Chris Frith of University College London and his colleagues found that when a person played the game known as paper, scissors, stone, another area associated with attributing intention, the anterior paracingulate cortex, fired more strongly if the subject thought they were playing a person rather than a computer — even if he or she was in fact playing a computer². What we perceive can depend on what we believe; imagining a thing produces brain activity very similar to the genuine experience.

As well as tapping into pre-existing biases,

film and animation mould the brain, says Frith. "The brain responds to cultural effects, and the conventions of realism are constantly changing," he says. What counts as real, according to these conventions, depends on technologies of representation; as technology has changed, so has our perception of the hallmarks of reality. People once believed newsreels only if they were in black and white. Currently, the mark of authenticity — exploited, for example, in Paul Greengrass's *Bloody Sunday* and *United 93* — is wobbly handheld footage. The coarseness of video stock, as used in *The Blair Witch Project*, is a similar signifier of reality.

Our understanding of the social brain is still rudimentary, Frith adds. "We know quite a lot about which regions are involved, but almost nothing about what they're actually doing, and

"People believe everything, and one must expend effort to disbelieve."
— Richard Gerrig

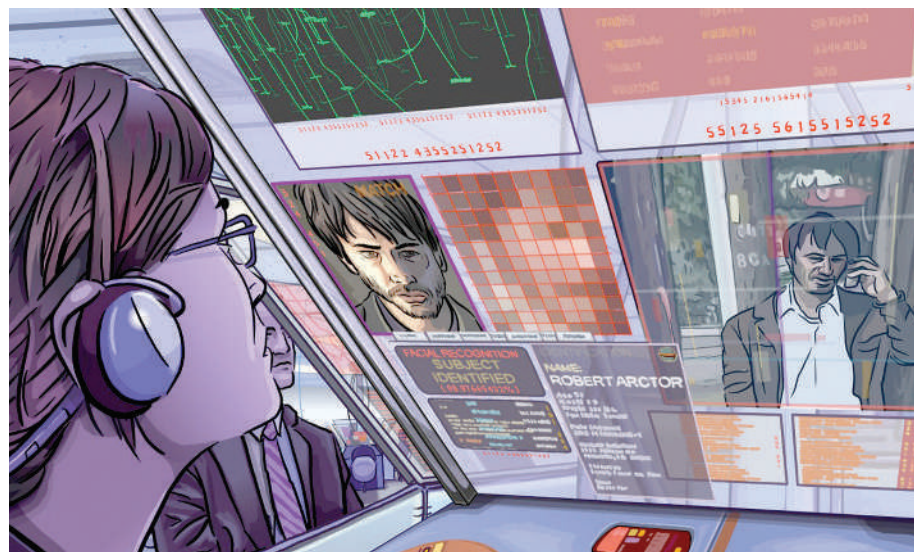
what neuronal computations are involved." What's needed, he says, are imaging studies that can reveal the timing of activity more precisely, showing how different brain regions interact, and theoretical analyses of how the brain solves social problems. "Ultimately, in a science-fiction way, I'd like robots that can do theory of mind and attribute intention." This would be beyond what even Dick imagined — in his novel *Do Androids Dream of Electric Sheep?*, filmed as *Blade Runner*, the Voight-Kampff empathy test is a way of spotting replicants' emotional deficiencies.

Easily deluded

The mental state that arises when we interact with unreality is complex. We get involved to the extent that, say, we cry when Bambi's mother dies, but not so involved that we walk out of the cinema and strike up a conversation with the nearest rabbit. Whatever the explanation is, says psychologist Richard Gerrig of the State University of New York, Stony Brook, it isn't the much-touted suspension of disbelief,

because disbelief is not the default. "People believe everything, and one must expend effort to disbelieve," says Gerrig.

The brain, it seems, has a default setting of credulity, and a keen appetite for consuming and producing stories. Narrative is a crucial tool in our efforts to understand the world and some brain areas seem specialized for processing it³. Information presented in narrative form lowers our critical faculties, and experiments show that the more deeply people become immersed in a story, the easier it is to sway their attitudes towards those advocated in that story⁴. This resonates with *A Scanner Darkly*, when an undercover cop becomes so engrossed in his 'fictional' identity as a drug dealer that his police persona begins to pursue his criminal one. Resisting our



Mind twisters: cartoon animation and real video footage seem to activate different brain areas.

susceptibility to stories is a useful skill in a media- and advertising-saturated world, says Gerrig. "We need to get kids and adults to construct disbelief. Because people don't know about this tendency, it puts them at risk."

"It's not important whether you label something as fiction or non-fiction," Mar agrees. "The true distinction is between narrative and non-narrative expository forms that don't draw you into their world." It also looks as if the abil-

ity to lose yourself in a fictional world might reflect your ability to navigate the genuine social world. Mar and his colleagues have found that the more time a person spends reading fiction the greater his or her empathy and social skills; for readers of expository non-fiction (such as, to pick an example at random, science journalism) the correlation is negative⁵. I thought it would be best to keep back that particular piece of reality until the end. ■

John Whitfield is a freelance writer based in London.

1. Mar, R. A., Kelley, W. M., Heatherton, T. F. & Macrae, C. N. (in the press).
2. Gallagher, H. L., Jack, A. I., Roepstorff, A. & Frith, C. D. *NeuroImage* **16**, 814–821 (2002).
3. Mar, R. A. *Neuropsychologia* **42**, 1414–1434 (2004).
4. Gerrig, R. J. & Rapp, D. N. *Poetics Today* **25**, 265–281 (2004).
5. Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J. & Peterson, J. B. *J. Res. Pers.* doi:10.1016/j.jrp.2005.08.002 (2005).

OLAF UNIV./COOP99/ JUICY FILM



In Benjamin Heisenberg's first feature film, a molecular biologist (left) informs on his colleague (centre, left) to a woman from the secret service, with chilling consequences.

Betrayal at the bench

Johannes walks across the grass to a Munich research institute on his first day in the lab. A woman approaches him. She is from the secret service and she wants him to spy on a fellow postdoc, Farid, a French citizen of Algerian origin, vaguely suspected of being a sleeper terrorist. Disgusted, Johannes refuses.

Johannes reports to his new boss, virologist Professor Behringer, who introduces him to Farid. Behringer, an authoritarian department head, wants the two young scientists to work on the same problem in molecular biology from different angles. Farid's approach is genomics, while Johannes is studying proteins.

Through work, Farid and Johannes become friends — of sorts. And therein lies a tale of trust, ambition and betrayal. By the end of *Schläfer* (Sleeper), a many-layered and earnest film selected for the 2005 Cannes film festival and now playing in European cinemas, Johannes will have informed on Farid. In doing so, he will have benefited his own career and may have won the girl.

Critics have praised this award-winning

first feature by a graduate of the Munich Film School. As an art-house film addressing big moral issues, it will not be to everyone's taste. But for an audience of scientists, three things will be remarkable.

First, that the director chose to set a study of post-9/11 paranoia in a molecular-biology laboratory. Second, that the laboratory setting, and the interactions between the scientists, are unusually realistic — even though the plot itself has nothing to do with science. And third, that the director is the grandson of Werner Heisenberg, 1932 Nobel laureate and originator of the uncertainty principle in quantum mechanics.

Benjamin Heisenberg was born in 1974, about a year before his grandfather died, but he is keenly aware of his forebear's legacy. The German physicist was criticized for working for the Nazi nuclear programme during the Second World War. But

Benjamin suspends judgement on Werner's decision to stay in Germany, asking: "How can anyone ever be sure that things turn out the way they expect and want, and that other generations won't judge their decisions differently?"

Werner was not the only high-achiever in the family. Benjamin's father Martin, a professor at the University of Würzburg, is one of the most highly cited behavioural scientists in Germany. His maternal grandmother was sister to Carl Friedrich von Weizsäcker, physicist turned philosopher, best known for his theories of the nuclear processes inside stars, and to Richard von Weizsäcker, the popular former German president. Two of his brothers studied sciences.

But Benjamin's interests took him to art school, where his interest in film awakened.

Film school was disappointingly superficial. "The philosophy of the school was to learn

"I realized how similar the procedures of film-making are to doing a research project." — Katerina zu Eulenburg

A. ABBOTT

craft, not art," he says, recalling the scepticism some teachers had of his earlier short films. "Their film history started with *Terminator II* and they told me 'oh no, don't do art, no one will watch it; it's just a mindfuck.'" He laughs at his own audacity and earnestness.

The idea for *Schläfer* came to Heisenberg shortly after 11 September 2001. "I saw that domestic security was being tightened and that no one was objecting," he says. "There would have been mass demos if politicians had tried to pass even a part of the new security laws ten years ago." Fear, he noted, was changing people's politics. It led him to wonder: how do fear and politics affect personal lives and relationships? How much does it take to weaken someone's moral convictions?

Growing up among scientists, Benjamin knew that the scientific environment had the dramatic potential for examining these questions. Scientists are dedicated, driven and sometimes ruthlessly competitive. Scientific conflicts could illustrate how much pressure — or how little — it takes to break a person's ethical spine. The international nature of science was also important to the plot, because an Arabic researcher would not be so unusual. Benjamin created a potentially explosive, but plausible, mix of characters: a ruthlessly ambitious principal investigator; and two equally ambitious, inexperienced young researchers who collaborate, yet compete with each other.

Two events test their friendship and collegiality. First, they fall in love with the same girl, Beate. When she chooses Farid, Johannes makes contact with the secret service agent he had previously spurned. But when Farid turns moody on Beate — he has learnt that he is being watched, but doesn't know by whom — she takes temporary comfort with Johannes. Uplifted by his romantic triumph, Johannes tells the secret service he will no longer speak with them.

Then a lab drama sets things on their final, fateful course. One day Farid barges excitedly into Johannes' lab: "*Nature* has said 'yes!'" Suppressing his annoyance that he didn't know a paper had been submitted, Johannes is thrilled. He had, after all, helped Farid's project by reanalysing his sequence data and identifying two overlooked genes that were key to solving their biological puzzle. They celebrate. But then he learns that the manipulative Behringer has excluded him from the author list. His rage is ignited, with chilling consequences for Farid. After Farid is arrested for suspected involvement in a failed Munich bombing, Behringer replaces Farid's name with Johannes' on the *Nature* paper.



Benjamin Heisenberg's film *Sleeper* asks what it takes to weaken a person's moral convictions.

We are never told the details of the research project. "People would stop listening — they only need to know enough to understand why a conflict has developed, and glimpse its complexity," says Benjamin. But any milieu has to be convincing to make viewers believe in the film. This is why Benjamin took pains to ensure the science throughout was as real as possible.

Enter Katarina zu Eulenburg, Benjamin's cousin, and an immunology PhD student at Berlin's Humboldt University. She read the script and hosted the actors in her lab, teaching them how to pipette liquids and handle animals. She was on set, ensuring that in each lab scene the actors and extras were working appropriately. She even provided her gloved hands for close-ups of detailed procedures. "It was such fun — and I realized how similar the procedures of film-making are to doing a research project," she says. "It is hard and disciplined work which you really have to believe in, because it takes such a long time to get the idea, write the script, find the financing and then make the film."

Her and Benjamin's efforts were rewarded when *Schläfer* won the 2005 Midas Prize for the best European drama featuring science. But after seeing the film, Martin Heisenberg told his son that no one — not even a scientist as ruthless as Behringer — can swap names on a paper already accepted by *Nature*. But Benjamin didn't reshoot. "The dramatic moment was too important to the plot,

and I was sure that one incorrect detail would not disturb the realism of the lab scenes."

His father suspects he was the source of Benjamin's view of the importance of a *Nature* paper, something he finds a little embarrassing: "I always tell my students that these things shouldn't matter so much." And Benjamin says he himself has only experienced labs with very positive atmospheres. "My portrayal of a pushy lab, whose competitive atmosphere became poisonous, comes from what other scientists tell me exists."

In *Schläfer*, this poisonous atmosphere leads Johannes to betray Farid, a decision we condemn but understand. His betrayal is foreshadowed by a scene in which Johannes kills a lab mouse. "We see a parallel in Johannes' ability to coldly kill an animal, even though we know from other scenes that he is a caring and sensitive person," says Benjamin. "Killing an animal for the greater good of science requires a decision to cross an ethical border — and then you just do it without thinking about the ethics every day." Benjamin imagines something similar must happen when someone decides to spy for the secret service.

Would the German secret service really recruit scientists in a public research lab? It's not out of the question, says Martin Heisenberg, who witnessed an east German spy in his former lab at the institute for virology at the University of Tübingen before German reunification. "He regularly went through notebooks and transmitted information."

Alexander Kekulé, a microbiologist at the University of Halle, Germany, says that spying on a suspected terrorist in a German lab today is entirely conceivable. "There have been recent scandals in Germany about the BND [secret service] getting journalists to spy on their colleagues — there would be even fewer scruples in spying in other professions, including science."

With *Schläfer*, Benjamin says he has got the big moral issues out of his system, at least temporarily. His next film is about a bank robber in Vienna.

But he knows he will come back to science in the future. "Mostly science in films is characterized by the classical mad professor, or someone running around saying 'Oh my God! The organ emulator is running at 400%!' — with no one having a clue what that could mean." But film-makers are becoming much more aware of the dramatic possibilities of science, he says: "The time of science in films is coming."

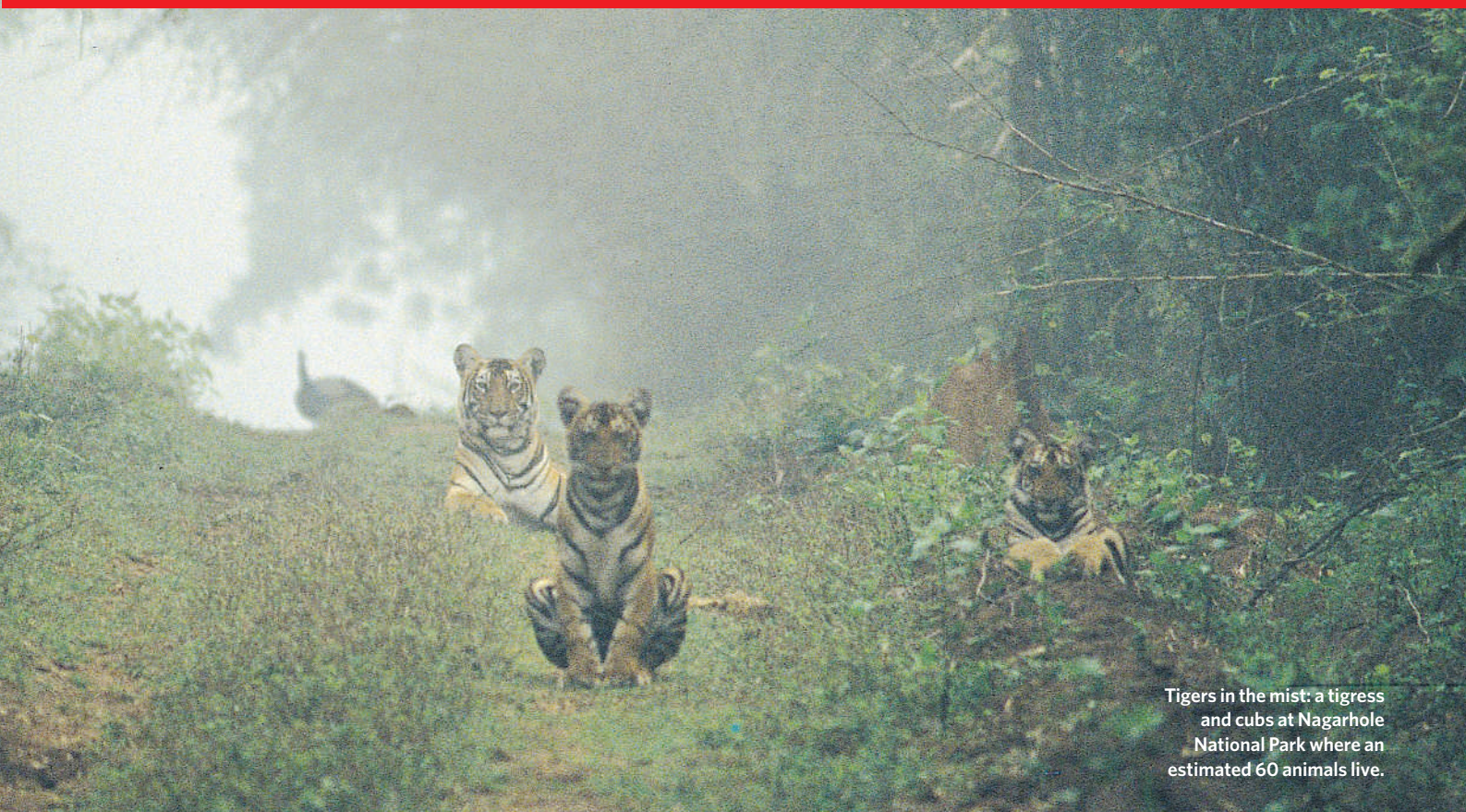
Alison Abbott is *Nature's* senior European correspondent.

The DVD of *Schläfer* with English subtitles is available from September. It can be bought from www.filmgalerie451.de.



When Johannes decides to spy on his colleague, he crosses an ethical line.

OLAF UNIV/COOP99/JUICY FILM



Tigers in the mist: a tigress and cubs at Nagarhole National Park where an estimated 60 animals live.

The tiger's retreat

U. KARANTH

Tigers are teetering on the verge of extinction and human contact in their habitat could be their greatest threat. **Erika Check** investigates whether local people can live alongside India's big cats.

It is the dry season in southern India, but a thunderstorm has cooled the air and damped down the dust in the forest of Nagarhole National Park in Karnataka. As he drives slowly down a dirt road in the park, wildlife biologist Ullas Karanth keeps one eye on the horizon and the other on the wet ground, scanning for tiger tracks.

Through a thicket of bamboo, we spot a young elephant ambling across a marsh. Karanth stops the jeep. The only sounds are barbets cooing from the trees and the frenzied cry of a cuckoo nicknamed the 'brain-fever bird'. Karanth murmurs, "This used to be a bustling little village... 30, 40 houses, lots of cultivation..." He doesn't finish his sentence, but his meaning is clear: now the people are gone, the wildlife is coming home to where it belongs.

We drive on, scanning the underbrush for any movement. Suddenly, Karanth hits the brakes and exclaims in an excited whisper: "There!" He points straight ahead, but at first I don't see it. A herd of spotted deer stands alerted by the road, about 20 metres ahead. I follow their eyes, and that's when I see him: the world's largest cat, pacing across the road.

I yelp; Karanth shushes me. But the tiger

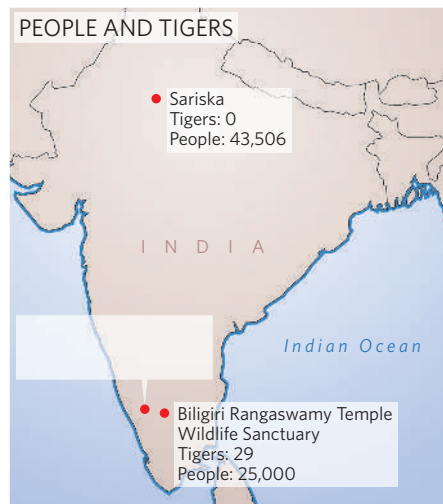
continues at his leisurely pace through the grassy clearing. I catch only a glimpse of his massive head and striped flanks before he disappears into the dense lantana and the woods beyond. The deer advance nervously after him, checking to make sure he's gone. Karanth shakes my hand. "You are really lucky," he says. "I have seen tigers so many

times, but each time it is like a special event."

For Karanth, each sighting is doubly satisfying, because his work in Nagarhole is part of the reason why tigers still exist there today. Karanth was the first Indian scientist to do rigorous tiger ecology in India. During his 20 years of work, he combined research with relentless activism to keep people and poachers away. And tigers have thrived there.

In the rest of India, the big cat has not been so lucky. A century ago, 40,000 tigers roamed India's forests and grasslands. Now, the Indian government estimates that only 3,600 still exist — and most informed observers say the number is probably between 1,200 and 2,000 (see 'Counting cats', overleaf). India's tiger population is teetering on the brink of extinction.

This is a tragedy, and not only for the tiger. The big cat is a source of national pride in India and a part of the country's identity. The tiger is also an umbrella species — a high-profile animal that earns protection for the other animals that live alongside it. This means that the same forces that threaten the tiger also bring bad news for many other vulnerable species in India, such as the rare Indian giant squirrel and the massive forest ox, the gaur.



COUNTING CATS

How many tigers are left in India? That depends on whom you ask. Since 1966, the Indian government has used the 'pugmark census' to track the beasts. Every few years, forest officials roam through every national park for one to two weeks looking for pawprints. They assign prints to individual tigers to arrive at a total count. The government's most recent estimate, four years ago, was 3,642 animals.

But biologists heap scorn on the pugmark census, saying it is virtually impossible to find tracks from every tiger in such a short time, or to tell the animals apart by their prints³.

Instead, most scientists have argued that the government should use sampling methods and statistics to monitor tiger populations.

As recently as 2004, Project Tiger director Rajesh Gopal flatly dismissed critics of the pugmark method and slammed the alternative — photographing and counting the animals with remotely triggered cameras. But after the embarrassing revelation last year that there were no tigers in Sariska National Park, Rajasthan — despite the government's claim that there were between 16 and 18 tigers there in 2004 — Project Tiger was forced

to make concessions.

Project Tiger has devised a monitoring strategy that involves both sampling and counting methods, and said last year that it would have an accurate estimate of India's tiger population by this July. But even the revised strategy has drawn heavy fire from international tiger biologists. Last month, Project Tiger relented under their criticism, halted its count temporarily, and said it would take at least another year to issue its final report. In the meantime, some conservationists think India's tiger population could be as low as 1,200 animals. **E.C.**

The reasons behind the tiger's decline are complex: rampant poaching; a lack of funding for the country's network of tiger reserves; a sclerotic and bureaucratic system of park management; the constant pressure of more people; and more development, such as industrial-scale mining in forests. Some scientists suggest that the government's distrust of conservation science is also behind the decline.

Tribal trouble

The conflict has led to emotional, seemingly intractable debates over the central question of tiger conservation in India: can tigers and people live together? Some, like environmentalist Sunita Narain, who led a government task force on tigers last year, say there is no choice. Today, 380,535 people live in India's 37,761 square kilometres of tiger reserves. Tigers and poor people live in the same spots, and both have rights to the land. But Karanth and other die-hard conservationists argue that tigers and people will never coexist peacefully.

Faced with this bleak picture, some have given up on the tiger's future. But not Karanth, who has watched the animals recover from seemingly hopeless odds in some places.

Karanth first visited Nagarhole in 1967, on a motorcycle trip during college. He hardly saw a deer, let alone a tiger. Logging and hunting were rampant. And India's tourism department was still luring visitors to the country by promising that they could shoot a tiger and take it home as an exotic souvenir.

But the animal's fate was not sealed. In 1972, Prime Minister Indira Gandhi convened a task force to study India's disappearing tigers. The

group's report expanded into the programme called Project Tiger. Today, that programme is responsible for the animal's survival in India, and controls a sprawling complex of 28 tiger reserves — from Namdhapa, on the eastern border with Burma, to Kalakad-Mundanthurai, at the southern tip in Tamil Nadu state.

Wildlife continued to fascinate Karanth as he pursued an engineering degree. After just three years working as an engineer, he realized he hated his career and changed course, leaving to do a PhD in wildlife biology at the University of Florida. He returned to Nagarhole in 1986 — this time, as a young scientist gathering data for a research project.

By that time, new park managers, led by forest officer K. M. Chinnappa, were working to stamp out poaching and enforce new wildlife protection laws. Nagarhole was starting to bounce back, and Karanth wanted to test its potential as tiger habitat. But undertaking a science project in the park was not easy. Chinnappa fought constant skirmishes with the 'tribals', groups of forest dwellers who lived by hunting, herding livestock and collecting firewood. Poachers, too, were angry that their prime hunting grounds were now off-limits.

Early on, Karanth and Chinnappa realized they shared a mission. Karanth had started a project to count tiger prey, from the tiny spotted deer to the hulking gaur. By figuring out how much tiger food was roaming around

Nagarhole, Karanth reasoned, he could predict how many tigers could live there.

But the studies were interrupted in 1992, when simmering tensions erupted into violence. An angry mob of villagers stormed the park, burned Chinnappa's house down, and vandalized Karanth's research station. A local man was murdered in the aftermath of the riot, and officials arrested Chinnappa for the crime. Although he was later cleared, Chinnappa was done with the forest department. He began working instead as a wildlife activist in the villages in and around Nagarhole, as well as in other national parks in the larger state of Karnataka. Karanth, too, realized that he could not do his science in isolation. Since then, he has worked with Chinnappa and other activists to ensure that Nagarhole and other parks remain untouched.

Numbers game

Today, Nagarhole has one of the healthiest tiger populations in India — about 60, by Karanth's estimation. He attributes that to both a crackdown on poaching and an effort to relocate 12 villages outside the park.

Karanth's belief that the tiger needs an inviolate habitat comes partly from his research. In the 1980s and early 1990s, Karanth perfected his method of counting prey to predict tiger populations, but absolute confirmation — by radio-collaring every tiger in the Nagarhole forests — wasn't feasible. So he decided to capture the animals another way: on film.

Karanth adapted a device called the Trailmaster, designed to help hunters find deer. It had an infrared sensor that, if tripped, triggers a camera to capture the passing animal's image. Karanth installed a handful of Trailmasters at Nagarhole and spent the next decade working with statisticians to estimate the tiger population based on the number of tripped sensors¹.

This technique, called camera-trapping, is widely used in wildlife biology.

Karanth's latest research, which combines ten years' worth of data, strengthens his conviction. In unpublished work, he estimates that over the past decade, an average of 23% of the park's tigers have died or left the park every year. And yet the population is stable, partly because females

breed often. To Karanth, this underlines the importance of keeping people out and sustaining prey populations for the tiger. "It's very resilient," he says. "But you can't have people, and you can't have cows."

Sweeping social and demographic shifts,

"There is no doubt in my mind that we need to do a lot to save the tiger — as much as we need to do to save poor people who live with the tiger."
— Sunita Narain



Candid camera: Ullas Karanth sets up camera traps to measure tiger numbers at Nagarhole.

however, are not in Karanth's favour. One billion people live in India, with a projected population of 1.6 billion by 2050. And the country's push towards social reform is empowering groups that have been traditionally disenfranchised.

In the past, tribal groups living alongside tigers were simply pushed off their land by the government. The forced relocations often left whole villages stranded with no way to make a living; villagers came right back to the forest. Now some scientists are asking whether there should be a different solution — some way for tigers and people to coexist.

Peaceful coexistence

Not far from Nagarhole lies a test case for coexistence, in the Biligiri Rangaswamy Temple Wildlife Sanctuary — BRT for short — in Karnataka. Inside the sanctuary are 540 square kilometres of rolling hills, rainforest, coffee plantations, and thousands of members of the Soliga tribal group. In 1972, when the sanctuary was created, the government moved 25,000 Soligas into small settlements called *podus* in and around the park. Now, they support themselves by harvesting lichens, fruits, honey and other forest products.

In one *podu* near the southern tip of the sanctuary, a group of Soliga women and children sits in a dirt yard circled by papaya, banana and mango trees. The women point out a shed with a dried-grass roof. Inside, stacks of burlap sacks bulge with lichens collected from the forests. It takes one week for a harvester to fill one sack, which earns the village 500 rupees — about US\$10. The harvesters pass the lichens on to a central cooperative that sells them and distributes the profits within the community.

Little evidence of profits can be seen in this *podu*. The houses are constructed of mud and sticks, and villagers must rebuild the grass roofs each year. The women wear tattered saris and are afraid to send their children to the school a few kilometres away because elephants might trample them on the way. A deep trench encloses the village; it's supposed to keep the elephants from raiding the crops. But if elephants don't get to the crops, marauding packs of wild boars do.

There is a new hospital an hour's drive away. But no one in the village owns a car, and they can't afford the 30-rupee bus fare. So what happens when someone gets sick? "We just die," says one Soliga woman, with a bitter laugh.

And yet the Soliga do not want to leave this *podu*. The state government has offered to resettle families on 300 hectares of land outside the sanctuary. But the men and women in this *podu* say they won't go. They are attached

E. BRIGGS

to their god, who lives in a hill that looms over the village. And they won't leave their temple, a spare, dirt-floored hut. "They have relationships with gods here, and they have burial sites within the sanctuary," says C. Madhegowda, an activist. Madhegowda, who is also a Soliga, explains: "We are children of the forest. We are not interested in leaving."

Meeting halfway

Madhegowda works with a group that is standing up to the government in ways that might have been unthinkable a few decades ago. In 2003, India banned the harvest of non-timber forest products from wildlife sanctuaries. In the past, the Soligas might have just accepted this removal of their main source of income. But not now. Madhegowda's group organized strikes in the Karnataka state capital, Bangalore, and met with officials in the forest department. For now, there is an informal détente that allows harvesting to continue in the sanctuary. "Twenty-five thousand people depend on this harvest," Madhegowda says. "You can't ban that if you don't provide an alternative."

The Soligas say they should be allowed to stay because they're not harming the park. Over the past decade, a Bangalore-based organization called the Ashoka Trust for Research in Ecology and the Environment (ATREE) has been working with the Soligas on sustainable harvesting practices. ATREE researchers have reported that the Soligas make accurate estimates of their harvest, and rarely use harmful methods, such as cutting branches off fruit trees. As a result, populations of harvested plants and bees have been relatively stable over the past decade².

ATREE researchers hope this will convince the state forest department to allow the Soliga to help manage the wildlife sanctuary where they live. "We're saying these people can take responsibility for managing this park," says Nitin Rai, an ecologist and sociologist with ATREE.

Rai started out as a 'wildlifer' like Karanth, but now believes it is not realistic to manage tiger populations by moving people out of their way. He points to the antagonistic relationships that have sprung up around some of the tiger reserves, where villagers are so alienated from the park's mission that they happily serve as paid guides for poachers. Such antagonism notoriously led to the poaching of every tiger from a reserve called Sariska in Rajasthan. The revelation last year that all Sariska's tigers were gone led to a public outcry,

Big cat country:
tourists on elephant
back can watch tigers
at close proximity
in some of India's
tiger reserves.



F. SAVIGNY/NATUREPL.COM

and the government's appointment of a task force to examine India's tiger crisis. "We've got to convince ourselves that it's OK to use management systems that include people as part of the ecosystem," Rai says.

A year-and-a-half ago, Sunita Narain would have taken Rai's side against Karanth. The outspoken activist has taken on some of the most powerful interests in India in her fight for environmental justice. Last year, her high profile and political connections earned her the job of chairing the prime minister's task force on Project Tiger. Narain's appointment was controversial, because she was an outsider to the tiger community.

Narain's unenviable task was to find some middle ground between the two camps of the tiger debate. And she believes the group did find a good compromise. Its final report estimated it would cost between US\$713 million and US\$2.5 billion to move every person out of India's tiger reserves. The task force instead suggested the government pick the most crucial areas of habitat and concentrate on moving people away from those places first. In the rest of the areas, the task force suggested that the government find ways for people and tigers to coexist, as some say they do in the Biligiri Rangaswamy Hills.

For Narain, this was a huge concession. "Coming into this as an environmentalist, I definitely thought that relocation was not the solution," she says. "But we tried to meet the conservationists halfway." And she stresses that the relocations, if they are done, should be done well. The villagers should be given good land, and enough money to make an honest go at a new life. This is a sore point in India, where government-run relocations are often

botched, leaving people in worse circumstances than those they left behind.

Feasible futures

On that point, Karanth agrees with Narain. The day after our tiger sighting in Nagarhole, we are up before dawn, driving to the forest-department quarters to meet a park official. There Karanth lays out a stern directive: he is trying to work out a deal to move one of the last villages out of Nagarhole — and he wants the forest department on his side. "Yes, yes," the forest official agrees, nodding as Karanth spells out his plans. But this is only one step, and many months of negotiations will follow.

Later, Karanth drives us past some of the villages that were relocated years ago to the outskirts of Nagarhole. There are neat rows of houses, and fields where men drive ploughs pulled by cows. Karanth sees this as inevitable progress. After all, he asks, should people live as the Soligas do forever — with no electricity, no hospital and no schools?

Some of the voices in the tiger debate disagree with Karanth's methods. But most now agree with his premise: that saving the Indian tiger also means helping the people who live alongside it. "There is no doubt in my mind that we need to do a lot to save the tiger — as much as we need to do to save poor people who live with the tiger," Narain says. "The poverty of each is leading to the poverty of the other."

Where possible, this will mean Karanth's strategy of relocating people to improve their welfare. Elsewhere, scientists, activists and officials must find other ways to end the antagonism between tigers and people. Otherwise, both species face a desperate future. ■

Erika Check is Nature's biomedical correspondent.

"We've got to convince ourselves that it's OK to use management systems that include people as part of the ecosystem."
— Nitin Rai

1. Karanth, K. U., Nichols, J. D., Kumar, N. S., Link, W. A. & Hines, J. E. *Proc. Natl Acad. Sci. USA* **101**, 4854–4858 (2004).
2. Shanker, K., Hiremath, A. & Bawa, K. *PLoS Biol.* **3**, e394 (2005).
3. Karanth, K. U. *et al. Anim. Behav.* **6**, 141–146 (2003).

BUSINESS



ICONOTEC/ALAMY; M. RIETSCHEL/AP/EMPICS

The old and the new: the ancient city of Dresden now houses one of Europe's most dynamic electronics hubs.

Angling Saxons

Eastern Germany is landing major electronics industry investments — but needs to build up its own innovative capacity, reports Ned Stafford.

Back in December 2000, when business leaders in the German state of Saxony founded an organization to boost their developing electronics industry, they chose to aim high. They christened it Silicon Saxony, after California's Silicon Valley — the high-tech champion of the world.

Nearly five and half years later, Silicon Saxony's boosters are still talking big. Helga Pratap, microelectronics expert at the State of Saxony's Economic Development Corporation, says the state's capital city of Dresden and surrounding area is now the address of choice for global firms expanding in Europe. "We are, by far, the most interesting cluster in microelectronics in Europe," she says.

Until a few weeks ago, sceptical observers might have dismissed Pratap's exuberance. But on 29 May, California-based chip maker AMD gave Dresden a huge stamp of approval, announcing that it would invest an additional US\$2.5 billion over the coming three years to expand production there. The investment will add about 420 new jobs to AMD's current workforce of 2,800, and bring its total investment to a spectacular \$8 billion since it first opened shop in Dresden in 1996.

Silicon Saxony now has 204 member companies employing 17,000 people, and generating estimated annual sales worth €3 billion (US\$3.8 billion). In addition to AMD, other large firms in the region include semiconductor maker Infineon Technologies, which was spun off from Siemens in 1999 and employs 5,500 people there, and Texas lithography

company Toppan Photomasks.

"We have achieved critical mass in Saxony," claims Peter Kücher, director of the Fraunhofer Center of Nanoelectronic Technology (CNT), a Dresden-based branch of the German network of applied research institutes known as Fraunhofer centres.

But Dietmar Edler of Berlin's German Institute for Economic Research says that unlike Silicon Valley, where venture capital plays an important role, most investment in Dresden is coming from just two big private companies — AMD and Infineon — and from government subsidies. "If you think it is necessary to go beyond production technology into creating new ideas and concepts," he says, "then more venture capital would be necessary."

Local knowledge

Nonetheless, Edler says, the region has progressed rapidly over the past decade. AMD and Infineon started with only factories, but have since added strong research and development capability, while the state and federal governments have encouraged and funded research at private and public institutions. "It is important to bind investors to the location and this has been done in Saxony in a step-by-step process," he says.

The Fraunhofer CNT is one of many public research laboratories in the region working closely with business to advance the technology behind microelectronics and the manufacturing process. The CNT — founded last year with the backing of Fraunhofer, Infineon, AMD, the

federal government and Saxony — is housed at one of Infineon's factories. The state has plenty of other research activity at seven universities, 12 Fraunhofer centres and six Max Planck Institutes doing basic science.

Daryl Ostrander, AMD's US-based vice-president of manufacturing, says that Dresden offers "world-class engineering" as well as an "economically friendly" environment. AMD, he says, utilizes "the considerable resources of the Fraunhofer and Max-Planck Institutes".

At first glance, Saxony might seem an unlikely home for an electronics cluster. Until the collapse of the Eastern Bloc in 1989, it was part of the old East Germany, and had no direct access to technologies developed in the West. The state currently has an unemployment rate of 17%, and an aggressive subsidy programme to lure investments such as AMD's.

Ulrich Schlicht, head of business development at the state's economics ministry, won't say how much of the \$2.5-billion investment was paid for by such subsidies. But he says that since 1990, the state has made grants of €1.1 billion to companies, supporting total investments worth almost €10 billion.

The state government has also spent about €300 million on electronics research programmes over the past 15 years, contributing, for example, to the Fraunhofer CNT.

Gerit Vogt of the Dresden office of the Munich-based Ifo Institute for Economic Research says that the links between local innovation and new chip factories are developing well in Saxony. Subsidies were needed to secure the initial investment, in his view. As a result, "there is more wealth, more income, more jobs," he says. "This is a story of success and I think it is a durable success." ■

Misconduct: lack of action provokes web accusations

SIR — As the webmaster of New Threads (www.xys.org) — the website “at the centre of concerns over claims of misconduct”, according to your Special Report “Named and shamed” (*Nature* **441**, 392–393; 2006) — I cannot agree with your comment that “some fear persecution reminiscent of that used in the Cultural Revolution”.

The Cultural Revolution was started by Chairman Mao in 1966 and formally ended with his death in 1976. Although 30 years have passed, the memory of this calamity is still vivid in many Chinese minds — it is understandable that some fear the tragedy might someday recur. But it is ridiculous to compare free speech on the Internet to the violence of the Cultural Revolution, which was controlled by a dictator, allowed for no freedom and included governmental persecution of ‘class enemies’. I find it ironic that 120 Chinese-American scientists and self-appointed human-rights advocates have signed an open letter appealing to the Chinese government to suppress media and public opinions: they still need to learn what free speech and human rights mean.

I agree that China should establish an official channel to investigate allegations of misconduct. In fact, I made this suggestion as early as 2001, in a speech to the Chinese students and scholars association at the University of California, San Diego (see www.xys.org/xys/netters/Fang-Zhouzi/science/yanjiang.txt). But before this channel exists, and to make sure it functions properly after it is established, free press and free speech are indispensable.

Shi-min Fang

New Threads Chinese Cultural Society,
PO Box 26194, San Diego, California 92196, USA

Misconduct: exposure is not like Cultural Revolution

SIR — Your Special Report (*Nature* **441**, 392–393; 2006) and Editorial “Finding fraud in China” (*Nature* **441**, 549–550; 2006) express deep concern about accusations of scientific misconduct in China. You rightly point out that it should be the government’s greatest priority to crack down on scientific misconduct, if it is rife.

The New Threads website covers wide areas such as literature and popular science. It is well known for posting accusations of all types of scientific misconduct, and providing a forum for people to discuss their concerns. There are good reasons for the popularity of the website among intellectuals and the general public. It is the motivation of those condemning it that needs to be questioned.

It is misleading to suggest that high-profile researchers could be persecuted through accusations made against them on the Internet, or to compare this to the Cultural Revolution. I witnessed the violence of the Cultural Revolution in my childhood. My parents were abused by the Red Guard because of their family, education and professional background. I cried when I saw crosses marking their names in posters and cartoons.

The Cultural Revolution was a mass movement organized by the country’s leader to crack down on his opponents. New Threads is just a platform without any official power: openness is the key to its success. It has become a portal for the grass roots who are ignored by official channels, such as university authorities, when they report misconduct. Internet debate and the resultant public attention can act as a warning to people attempting to violate research ethics. This is nothing like the horror of the Cultural Revolution.

Zheng Huang

Radiation Oncology Department, University of Colorado at Denver and Health Sciences Center, Aurora, Colorado 80045, USA

Misconduct: Chinese funding body unmoved

SIR — I can believe you had difficulties contacting the National Natural Science Foundation of China (NSFC) for your Special Report (*Nature* **441**, 392–393; 2006). As executive editor of the *Journal of Clinical Investigation*, I recently attempted to contact the NSFC to help us investigate an allegation of misconduct in a study the journal had accepted from Chinese authors.

A whistle-blower’s e-mail from someone at the authors’ institution indicated that data in the study were fabricated. The authors had collaborated with the dean of their university, so we were unsure whether contacting the institution would result in an unbiased investigation. As China does not have a supervisory body akin to the US Office of Research Integrity, we thought the NSFC — which had funded the study — would help.

Despite multiple e-mails, in English and in Chinese, the NSFC did not respond. Through a personal contact, I was put in touch with the director of the division that granted funding to the senior author. After repeated e-mails, I received the following: “I received the letter you written and we discussed the things you written to him. We have to say it is very difficult for us to determine whether their work is true. Because there are more than 50,000 proposals and about 10,000 grants supported in the Foundation every year. So we think maybe you have to find other way to make sure the thing.”

This statement indicates that the NSFC does not prioritize the policing of misconduct.

If it is not responsible for, or is too busy to investigate, the researchers it funds, then surely such behaviour is tolerated and endorsed?

The story has a sad ending. When confronted, the senior author claimed to have misplaced the primary data during a move between laboratories. In addition, the co-corresponding author (who signed our authorship agreement form upon acceptance, and who wrote to our office inquiring about the publication date) later claimed never to have seen more than the title of the work and asked to be removed as a co-author. The authors have now withdrawn their article.

Alongside Xin-Yuan Fu and his 120 co-signatories, I too eagerly await a response from Chinese authorities on whether they will establish a body to police misconduct.

Ushma Savla Neill

Journal of Clinical Investigation, Columbia University College of Physicians and Surgeons, Box 57A, New York, New York 10032, USA

Education and training put Iran ahead of richer states

SIR — We read with interest your News report “Arab state pours oil profits into science” (*Nature* **441**, 132–133; 2006). Other countries in the Gulf have also tried to spend oil money on setting up branches of Western universities, which is of arguable value when the infrastructure and the basic prerequisites of scientific research do not exist. Educating and training the personnel capable of doing research, as you describe in Qatar, is more important than spending on research and buying sophisticated equipment. Focusing on research and inviting scientists from overseas may lead to some short-term results, but it does not guarantee sustainable development without a solid, internal educational base.

Iran is a good example of a country that has made considerable advances through focusing on education and training. Despite sanctions in almost all aspects of research during the past 27 years, Persian scientists have been producing cutting-edge science. Their publication rate in international journals has quadrupled during the past decade. Although it is still low compared with the developed countries, this puts Iran in the first rank of Islamic countries.

Considering the country’s brain drain and its poor political relationship with the West, Iran’s scientific community remains productive, even while economic sanctions make it hard for universities to purchase equipment or send people to the United States to attend scientific meetings.

Mohammad Reza Mohebbi*, Mehri Mohebbi†

*Department of Neurology, University of Texas Medical School, Houston, Texas 77030, USA

†Center for Immigration Research, University of Houston, Houston, Texas 77204, USA

BOOKS & ARTS

Lessons from Italy

How malaria affected, and was controlled in, Italy in the past century has important messages for today.

The Conquest of Malaria: Italy, 1900–1962

by Frank M. Snowden

Yale University Press: 2006. 288 pp.

£25, \$40

Brian Greenwood

Much has been written about the biology of malaria but little about its social and economic impact. Frank Snowden has drawn on neglected Italian sources to produce a detailed account of the protracted struggle to bring malaria under control in Italy. *The Conquest of Malaria* is an important contribution to the malaria literature.

Snowden, a historian, uses contemporary accounts to illustrate the devastating social and economic consequences of malaria in Italy 100 years ago. On the eve of the First World War, life expectancy in malaria-endemic areas of Italy was only 22.5 years, less than in the least-developed countries today. Malaria was primarily a disease of the poor and was particularly severe among migrant agricultural workers and Sardinian miners who lived in appalling conditions.

Snowden points out that the detailed statistics available on numbers of cases of malaria and malaria deaths provide only a partial picture of the impact of the infection. It was widely recognized at the time that it had a major adverse effect on the economy, contributed to the impoverishment of southern Italy and was probably a contributor to emigration. Through both its direct and indirect effects, malaria may, at its peak, have caused as many as 100,000 deaths per year in Italy.

Malaria thrives during periods of turmoil, so it is not surprising that some of the gains in control of the disease made by the time of the First World War were lost during the period of the conflict. Troops and civilians involved directly in fighting in the Veneto region suffered badly, but malaria also surged elsewhere. Snowden attributes this to the return of infected soldiers to susceptible communities, shortages of antimalarial drugs, breakdown in rural health services and, less plausibly, to the 1918 influenza epidemic. During the Second World War, there was again an increase in cases of malaria for similar reasons, but with an additional, more sinister, explanation. Before retreating, the occupying German army flooded the Pontine Marshes, where malaria had previously been controlled, with salt water



Italians faced a malaria epidemic at the end of the Second World War after the German army withdrew.

to create the brackish conditions that favour the breeding of the mosquito *Anopheles labranchiae*, the main malaria vector in the area. This attempt at biological warfare was successful, resulting in an epidemic of malaria in 1944–46.

The description of the ways in which malaria control in Italy evolved over time provides many important messages for today. At the beginning of the twentieth century, quinine was seen as the main control tool. Its widespread use was highly successful in reducing malaria mortality but less efficacious in reducing the number of cases. This is not surprising as quinine does not prevent relapses of the malaria parasite *Plasmodium vivax* or kill gametocytes. There was resistance to its use in some areas because of suspicions over the intentions of employers and the government who provided it free. Provision of rural dispensaries, which provided general care as well as malaria treatment, and rural schools helped to overcome these concerns.

After the First World War, vector control started to receive more attention and became a key component of an integrated malaria control programme in the Pontine Marshes that was undertaken by the fascist government.

Snowden outlines the political motivation for this successful programme, which allowed 60,000 people to be settled in an area that had previously been largely uninhabitable. After the Second World War, the insecticide DDT became the primary control tool, and its use led to the elimination of local malaria transmission in Italy in 1962. However, this success was built on an enormous amount of background knowledge and experience in malaria control gained during the previous 60 years.

Could this excellent book have been improved in any way? Snowden uses specific events to illustrate general principles but it is sometimes difficult to know how these relate to the overall picture prevailing at that time, and a figure or two summarizing the chronology of events would have been useful. Some more maps would also have helped, for example to show regional differences in incidence rates. Snowden gives a good account of the biological aspects of the story and I found only a few, minor technical points with which I disagree. However, I think he is unfair to the World Health Organization (WHO) in implying an overreliance on the use of insecticide-treated bednets in current malaria control programmes; the WHO also advocates early

STUDIO PATELLANI/CORBIS

diagnosis, treatment with effective combination therapy and chemoprevention in pregnancy.

What lessons can be gained from the Italian experience that are relevant to malaria control today? There are many, but the main one, I think, is that malaria can be controlled successfully, even in the most extreme situations. What's needed is an integrated approach — rather than reliance on a single tool — based on sound epidemiological knowledge, supported

by an effective rural healthcare and educational programmes, and with a high level of political and financial support. These are demanding requirements, but the Italian story described in this important book shows what will be needed if malaria is to be controlled effectively in Africa today.

Brian Greenwood is at the London School of Hygiene and Tropical Medicine, London WC1B 3DP, UK.

adds: “Mistakes play a fundamental role in shaping the form of science — when they reveal themselves, they require new theories to explain why they are in fact mistakes. This gap between theory and reality provides the space for knowledge to develop.” These are central themes: science makes progress, but not directly or predictably; and science is done by people who have idiosyncratic quirks and convictions.

Neither of these themes is new to historians and philosophers of science, but they continue to surprise many scientists. Even scholars who have written extensively about the history of embryology (including myself) have not always done a good job of making these ideas available to a wider public, nor have they pulled together these particular episodes in one place. Therefore, Cobb's exuberant story is welcome. Much of the ground he covers is familiar only to professional historians of science, not to those working in the scientific traditions discussed here.

Cobb makes the science accessible. Who will not, for example, be delighted to picture the Pitti Palace in Florence as surrounded by wonders of “rotting flesh, buzzing flies and wriggling maggots” — ripe for studying whether flies come from other flies or spontaneously from the rot? Or to learn of debates and battles for primacy as Nicolaus Steno, Swammerdam and a cast of others jostle like schoolboys to establish who will be best friends with whom in a given year.

The book also does a wonderfully engaging job of presenting research on sperm, eggs and fertilization that drew on methods of dissection, observation and experimentation. Cobb's lively stories make the process of scientific discovery and adjudication approachable and intriguing. The title of the British edition is clearly intended to evoke fun-filled images of the egg-and-spoon races of school days and country fairs. Even though the book is not really about a defined ‘race’, and even though the seventeenth century actually brought only the very first steps in understanding generation, the book is welcome because Cobb has so much fun showing us how the story of sexual reproduction in animals began to be put together. His website (www.egg-and-sperm.com) contains a rich collection of additional illustrations and information.

My only quibble is that Cobb too easily evaluates past science as a ‘mistake’, even while acknowledging that it made sense in its context. Why consider it mistaken, then, rather than the best explanation in the circumstances? It is neither necessary nor helpful to judge science in this way. That said, Cobb's approach and enthusiasm are infectious. More history of science should be this enticing and accessible.

Jane Maienschein at the Center for Biology and Society in the School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501, USA. Her latest book is *Whose View of Life?*

Life under the microscope

The Egg and Sperm Race: The Seventeenth-Century Scientists Who Unravalled the Secrets of Sex, Life and Growth

by Matthew Cobb

Free Press: 2006. 352 pp. £17.99. Published in the US in August by Bloomsbury.

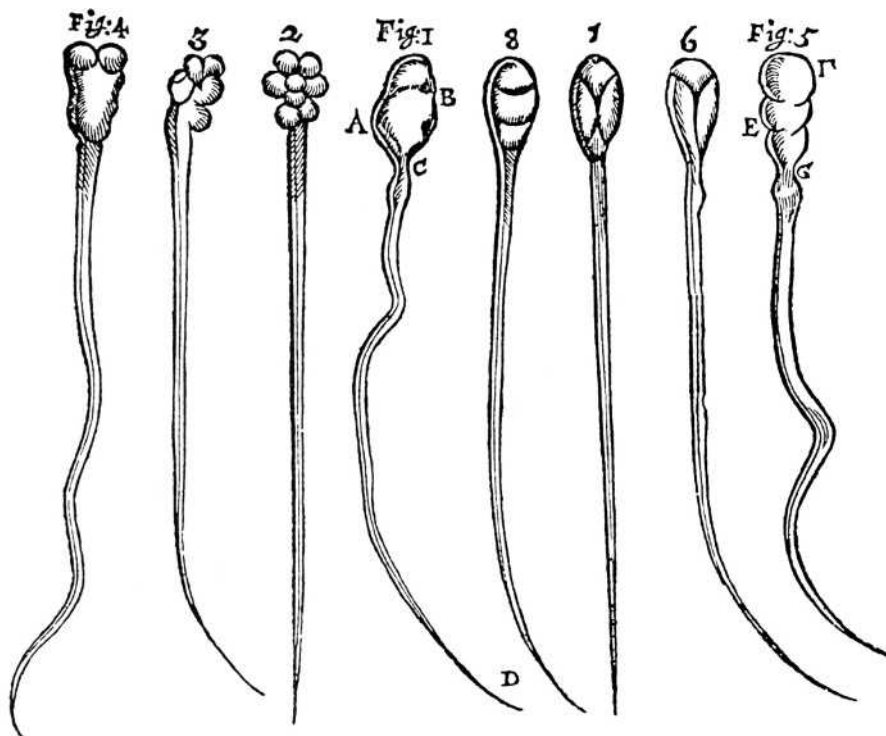
Jane Maienschein

A lecturer in the life sciences, Matthew Cobb is enchanted by science and its powers of discovery. He was studying insect behaviour and communication when he became intrigued by the seventeenth-century Dutch microscopist Jan Swammerdam. Disappointed that few people knew of Swammerdam's research, Cobb sought to make it better known. This book is the result.

The Egg and Sperm Race — or *Generation* as the US edition is called — is about seventeenth-century research on sperm, eggs and fertilization. That research primarily involved insects and a few other organisms, but not humans. After the early explorations of sex

and fertilization in other species, it took another 150 years to put together a coherent story about reproduction generally, and even longer to do more than surmise that humans are reproductively pretty much like other mammals. Only in 1828 did Karl Ernst von Baer first report seeing mammalian eggs, and only in the 1840s did researchers observe the merging of egg and sperm cells. Researchers worked out some of the processes of cell division, and by 1900 had just begun to understand the role of the nucleus and chromosomes in development. But it was not until the late 1970s, with *in vitro* fertilization, that researchers even began to be able to observe fertilization and cell division.

The story is about adventures in science, by researchers who were passionate about their work and at times highly competitive and even combative. Science is not a straight-line path to knowledge, Cobb writes, but rather “takes strange detours, finding itself temporarily trapped in unexpected dead ends”. He



Seminal work: sperm from rabbits and dogs, drawn by Antonie van Leeuwenhoek in 1678.

Virtually wild

Dazzling the millions who troop past, the habitat dioramas of New York's American Museum of Natural History, including this gorilla showcase, have fused science and art into hyper-real three-dimensional displays of animals in their native savannah, swamp or plain. *Windows on Nature: The Great Habitat Dioramas of the American Museum of Natural History* by Stephen Christopher Quinn (Harry N. Abrams, \$40) takes a behind-the-scenes look at more than 40 of them, each a testament to the explorers, naturalists, taxidermists and artists who painstakingly brought them to life more than half a century ago. This heroic task — it took 18 years to complete the 29 dioramas of the North American mammal section alone — inspired a number of conservation measures, as well as considerable awe. **B.K.**



D. FINNIN/AMNH

A weird, wired world

Entangled World: The Fascination of Quantum Information and Computation

edited by Jürgen Audretsch

John Wiley: 2006. 312 pp.

\$35, £22.50, €33.80

Vlatko Vedral

Quantum entanglement seems to be at the core of all the weirdness in quantum mechanics. Erwin Schrödinger called it the characteristic trait of quantum mechanics. Richard Feynman said that there was only one mystery in quantum mechanics and everything else can be seen as just its (frequently complex) consequences.

Entanglement — a bewildering set of correlations between objects in the quantum world — has historically been thought to lead to various physical, and possibly logical, paradoxes involving notions such as non-locality and cats that exist simultaneously in dead and alive states. Interestingly for engineers and computer scientists, all this quantum weirdness, far from being a hindrance, can be used to construct quantum computers that can run much faster than any current computer. Far from just being a theoretical quirk, small-scale quantum computers have been experimentally made that can perform simple calculations.

At least as interesting to physicists, although somewhat more controversial, is the issue that the very existence of the classical (non-quantum) world also crucially depends on quantum weirdness, or more specifically quantum entanglement. To put it more provocatively, the classical world behaves predictably, deterministically and in a localized way only because of the entanglement that exists on the larger scale in the Universe. This hypothesis

goes under the name of decoherence, which can be thought of as synonymous with the entanglement between a physical system and its environment.

In fact, quantum computation and decoherence are two different sides of the same coin. Entanglement that can be controlled is used in quantum computers; entanglement with the environment that is uncontrollable causes decoherence, destroying quantum computers and facilitating the emergence of the classical world. The distinction between controllable or uncontrollable entanglement makes no difference to the Universe, and is only important to the tool-building human mind.

Entangled World is a nice collection of essays and reviews addressing all three of the above aspects of entanglement: the philosophical (dead-and-alive cats), the physical (decoherence) and the technological (quantum computers). This makes it a unique collection, and is its strongest selling point.

The book contains several basic yet erudite expositions, many written by people who have made significant contributions to the field, such as Reinhard Werner, Rainer Blatt, Gerhard Rempe, Harald Weinfurter and Erich Joos. The first few essays are on the basic principles of quantum mechanics, and then the book moves on to entanglement and its use in quantum computation, teleportation and other quantum wizardry. The second half of the book contains reviews on the emergence of the classical world through decoherence, concluding with an essay by Michael Esfeld on what all this really means. Physicists like me, although always wary and suspicious of philosophy and the soundness of its reasoning, still like to read such musings about the meaning

of quantum physics. This is, after all, why we do science: to understand and explain the world we live in.

My only concern about the book is that it perhaps sits somewhat uneasily between being a 'popular science' book and a third-year undergraduate introduction to the field. Some of the early essays could easily be part of a popular book, containing as they do very few equations, whereas the reviews of quantum computation and decoherence are more technical and do require some undergraduate physics training, although nothing terribly excessive. More consistency of style and level throughout the book would have been desirable, so it would totally captivate at least one of its target audiences.

But what about some of the really fundamental questions in quantum physics? Can large-scale quantum computers be made that would operate at room temperature? Is the world really universally quantum mechanical, or do we also need the notion of a classical world? If we do, where exactly is the division between the two worlds? Can cats (or indeed humans) really be dead and alive at the same time? Are macroscopic superpositions just a question of money and experimental ingenuity, or will they one day collapse together with quantum physics? Unfortunately, the reader is unlikely to find answers to these questions in *Entangled World*, or indeed in any other physics book.

What you will find in *Entangled World* is a comprehensible, friendly and insightful introduction into modern ideas in quantum physics, including some fascinating, albeit brief, excursions into quantum gravity. These are the salient points to further enthuse you about the subject and perhaps motivate you to do a little thinking about the big questions. ■

Vlatko Vedral is in the Quantum Information Group, School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK.

NEWS & VIEWS

ANIMAL BEHAVIOUR

Trust in fish

Lee Alan Dugatkin

A mutually beneficial interaction between two species of fish turns out to involve the careful appraisal of one by the other — and the appropriately virtuous behaviour by the former while being watched.

Many animals, including humans, are score-keepers, paying close attention to what others do to them and responding in kind. This sort of tit-for-tat reciprocity works well, but is limited to instances involving some sort of direct, one-on-one, interaction. What, however, if organisms could learn about who was trustworthy (or not) by observing others from afar? Such 'eavesdropping', as Bshary and Grutter show on page 975 of this issue¹, characterizes an amazing case of cooperation through social networking in two species of fish: the cleaner wrasse and the bridled monocle bream.

Social scientists have long realized that humans are part of complex social networks, within which information travels in many directions. It took time, however, for this view of behaviour to migrate to other disciplines, particularly behavioural ecology. Although the sociobiologist Richard Alexander² discussed indirect ways of obtaining information back in 1987, only recently have evolutionary and behavioural ecologists begun to take the study of such social networks in non-humans seriously. This has occurred along two distinct, but parallel, paths.

One path towards understanding social networks centres on the evolution of cooperation. Here, tit-for-tat scorekeeping is expanded, such that individuals can glean information on the cooperative tendencies of others by observing behavioural interactions in their group. They then create what is called an 'image score' of those around them — this score increases when an observed individual cooperates, and otherwise decreases^{3,4}. Image scoring allows cooperators to preferentially help other cooperators, and sets the stage for the evolution of reputation⁵. Such a scoring system may also create selection pressures for individuals to advertise their cooperative tendencies⁶.

The second path for studying social networks involves eavesdropper effects and audience effects⁷. These effects often centre on territorial behaviour, and more generally aggression, in which eavesdroppers learn about the fighting abilities of those they monitor, and use this information when they



Figure 1 | Client and cleaner. The relationship between these two species, the bridled monocle bream and the cleaner wrasse, constitutes a complex social network.

subsequently interact with such individuals. In principle, eavesdropper effects, as well as audience effects — in which the individual being observed behaves differently as a consequence of being watched — are not restricted to aggressive behaviour, but can evolve in almost any context.

Bshary and Grutter¹ merge the image-scoring and eavesdropper approaches to uncover a fascinating example of a complex aquatic social network involving the cleaner wrasse (*Labroides dimidiatus*) and its 'client' bream (*Scolopsis bilineatus*) (Fig. 1). In this system, the wrasse 'cleans' the bream by eating the parasites that live on its body. The catch is that the cleaner wrasse actually prefers eating the client's mucus, and such cheating is not in the client's best interest. Yet, clients never eat their cleaners in Bshary and Grutter's experimental set-up — so what can clients do to prevent cheating, and why doesn't such cheating characterize this system? The answers are steeped in a social-networking system that has evolved between the two species.

To disentangle this network, Bshary and Grutter constructed an ingenious experiment in which a client fish could observe two clean-

ers. One of the cleaners appeared to be interacting cooperatively with another client fish, whereas no information about the cooperative tendencies of the second cleaner was provided. Given a choice between two such cleaners, the client preferred to spend its time near the cooperator. That is, clients chose their cleaners by eavesdropping on them, and preferentially interacting with those most likely to actually eat parasites, rather than with unknown individuals, who might very well try to get a quick meal of nutritious mucus off the client.

On the other side of the social network, Bshary and Grutter examined whether cleaner fish act in a more cooperative manner when they are being watched. Such audience effects would make cooperation all the more likely to evolve. Bshary and Grutter searched for such effects by setting up an artificial pulley system attached to a series of plates that contained two food items — flake food and prawns. Of these two items, cleaners prefer prawns over flake food (just as they prefer mucus over parasites on their clients). The plates, then, acted as proxies for the client fish: that is, the plates 'watched' the cleaner fish and adjusted what food was available as a function of the

A. GRUTTER

behaviour of the cleaners. By using the pulleys, Bshary and Grutter could test whether cleaner fish were more likely to take the less preferred food, if taking the more preferred food from a plate meant that they would lose access to other clients in the wild if they cheated). And indeed, the authors did find audience effects under these admittedly contrived conditions, as cleaners were induced into eating the less preferred foods.

If social networks such as those uncovered by Bshary and Grutter in fish are common in other animals, the prospect for researchers is both exciting and daunting. It is exciting because it suggests that animal behaviour has to be seen as being embedded in a rich social tapestry,

composed of interconnected direct and indirect streams of information. And it is logistically daunting for exactly the same reason. ■

Lee Alan Dugatkin is in the Department of Biology, University of Louisville, Louisville, Kentucky 40202, USA.

e-mail: lee.dugatkin@louisville.edu

1. Bshary, R. & Grutter, A. S. *Nature* **441**, 975–978 (2006).
2. Alexander, R. *The Biology of Moral Systems* (Aldine De Gruyter, New York, 1987).
3. Nowak, M. A. & Sigmund, K. *Nature* **393**, 573–577 (1998).
4. Wedekind, C. & Milinski, M. *Science* **288**, 850–852 (2000).
5. Pollock, G. & Dugatkin, L. A. *J. Theor. Biol.* **159**, 25–37 (1992).
6. Zahavi, A. & Zahavi, A. *The Handicap Principle* (Oxford Univ. Press, 1997).
7. McGregor, P. (ed.) *Animal Communication Networks* (Cambridge Univ. Press, 2005).

ASTROPHYSICS

Magnetic accretion

Daniel Proga

Disks of hot gas drawn onto a central star or black hole are the best energy-producing machines in the Universe. So how do these accretion disks work? The answer, it seems, is blowing in their winds.

Astronomical objects that siphon off their energy from gravitational processes are the most powerful sources of electromagnetic radiation in the Universe. Of these, quasi-stellar radio sources (quasars), which can release as much power as several hundred galaxies, are perhaps the most spectacular example. But for all their brightness, gravity-powered sources are secretive when it comes to revealing how exactly they produce their radiation. On page 953 of this issue, Miller *et al.*¹ investigate a system known as GRO J1655–40, in which a black hole of seven solar masses accretes gas stripped from a normal star. They observe not just radiation, but a wind of ionized gas emerging from the system, and conclude that a magnetic process powers this wind — and that this process could also play the lead part in producing the radiation.

For astronomers, gravity is not just the seemingly all-conquering force of everyday experience, but also the ultimate source of energy. In a star depleted of nuclear fuel, for instance, gravity prevails to form a very dense, compact object. Depending on the star's initial mass, this object could be a white dwarf, a neutron star or a black hole. When rotating gas is fed to such a compact object, as for example in GRO J1655–40, its mass is converted to radiation with an efficiency of a few per cent. That climbs to 40% for a rapidly rotating black hole — values significantly higher than the 1% conversion efficiency for nuclear fusion².

So how do these cosmic engines work? Initially, the falling gas forms a thin, dense 'accretion disk' that encircles the compact object and

is supported against gravitational collapse by centrifugal force. Through viscous damping processes within the disk, the mechanical energy of the gas is dissipated and radiated away as continuum emission over all wavelengths. That much is understood; what is unclear, however, is how exactly the mechanical energy is converted into heat and radiation energy. Moreover, for gas to accrete through the centrifugally supported disk, conservation of angular momentum requires a mechanism that transfers the gas's angular momentum rapidly outwards. Otherwise, the gas would orbit the central object forever, and no radiation would be produced.

For decades, hydrodynamic processes were thought to supply this mechanism. But despite some successes — such as obtaining an understanding of the role of turbulence in disk dynamics — the hydrodynamical approach has so far not produced a consistent model that can be derived from first principles^{3,4}. Only a turbulent disk with interacting eddies will supply an effective viscosity large enough to drive the fast accretion rates observed in disks; but the question of what powers the turbulence remains unanswered.

In this context, the introduction of magnetic fields, presumed to originate in the star feeding the compact object, pays off handsomely. This applies even to small fields that are hard to measure directly: through an effect known as magneto-rotational instability⁵, a weak magnetic field can act as a kind of spring connecting inner and outer areas of gas in the disk. This spring exerts a force that transfers angular

momentum and energy outwards and extracts disk rotational energy to power turbulence⁴.

Miller and colleagues¹ find indirect evidence to support this theory in the properties of a wind driven from the accretion disk of GRO J1655–40. They studied an X-ray spectrum of the system taken by NASA's Chandra satellite observatory and detected tens of spectral absorption lines shifted towards shorter wavelengths. Such a 'blueshift' indicates that absorbing gas is moving towards the observer and, using geometrical and physical arguments, Miller *et al.* conclude that the absorbing medium is a wind blowing outwards from the accretion disk at a small angle to the disk's plane. The fact that the wind is blowing from the disk and not, for example, from the companion star means that the wind's properties can reveal the properties of the disk itself.

To produce such a wind, the upper disk atmosphere must gain extra energy to overcome gravity. This extra energy can come from heating, from radiation pressure, from magnetic fields or from any combination of the three^{3,6,7}. In many systems, it is hard to distinguish which of these dominates. Happily, Miller *et al.* argue¹, the GRO J1655–40 case is clear-cut. In this system, the wind is too cold for its thermal pressure to overcome gravity, and the radiation pressure is too weak to drive a wind. Thus, by a process of elimination, a magnetic origin for the wind is the only option. This is an attractive possibility, as theories exist that posit magnetic fields as the drivers of disk accretion in the first place^{7,8}, and a larger mystery would therefore also be solved.

Not all might find Miller and colleagues' arguments for magnetic fields as the drivers of disk accretion convincing, partly because they rely on evidence that is mostly indirect. But this is the best that can be done at the moment: current models lack the predictive power to demonstrate that magnetic fields power the wind and disk accretion that are responsible, respectively, for the observed absorption lines and continuum emission. Nevertheless, many astronomers will be motivated by such work to develop and explore this promising model. Clearly, much can be learned about accretion disks by studying the properties of what escapes them — not only radiation, but gas, too. ■

Daniel Proga is in the Physics Department, University of Nevada Las Vegas, 4505 South Maryland Parkway, Las Vegas, Nevada 89154-4002, USA.

e-mail: dproga@physics.unlv.edu

1. Miller, J. M. *et al.* *Nature* **441**, 953–955 (2006).
2. Frank, J., King, A. & Raine, D. *Accretion Power in Astrophysics* (Cambridge Univ. Press, 2002).
3. Shakura, N. I. & Sunyaev, R. A. *Astron. Astrophys.* **24**, 337–355 (1973).
4. Balbus, S. A. & Hawley, J. F. *Rev. Mod. Phys.* **70**, 1–53 (1998).
5. Balbus, S. A. & Hawley, J. F. *Astrophys. J.* **376**, 214–233 (1991).
6. Begelman, M. C. *et al.* *Astrophys. J.* **271**, 70–89 (1983).
7. Blandford, R. D. & Payne, D. G. *Mon. Not. R. Astron. Soc.* **199**, 883–903 (1982).
8. Miller, K. A. & Stone, J. M. *Astrophys. J.* **534**, 398–419 (2000).

CELL BIOLOGY

The Golgi grows up

Vivek Malhotra and Satyajit Mayor

The Golgi apparatus of the cell has long baffled biologists, mainly because it is unclear how proteins are conveyed through it on their way to the cell surface. Some innovative microscopy may resolve the issue.

The Golgi apparatus is the enigmatic organelle responsible for modifying newly synthesized proteins that are destined to be secreted from the cell. Discovered by Camillo Golgi in 1898, the apparatus consists of a stack of disc-shaped membranes called cisternae. Newly made secretory proteins are delivered to the *cis* face of the Golgi. They exit, fully processed, at the opposite end — the *trans* face. But how do the proteins traverse the Golgi stack? Debate on this hotly contested question has persisted, in one form or another, for about 40 years. Now, two papers in this issue^{1,2} claim to resolve the matter in favour of one of the two competing transport mechanisms — the so-called ‘cisternal maturation model’.

This model proposes that the cisternae progress through the Golgi, gradually moving through the stack as new layers form at the *cis* face and old layers disperse from the *trans* face, and that they carry the secretory proteins with them (Fig. 1a). By contrast, the ‘vesicle-shuttle model’ (Fig. 1b) proposes that Golgi cisternae are long-lived structures, with secretory proteins being transported from layer to layer in small bubble-like membrane vesicles, which bud off one cisterna and fuse with the next one to disgorge their protein cargo.

At the heart of the controversy are the transport vesicles called COPI vesicles — after the coat-protein complex I (COPI) proteins that encrust their surface. There is a broad consensus that COPI vesicles bud from Golgi cisternae³, but no clear evidence about what these vesicles carry, or in which direction they travel — *cis* to *trans*, or the reverse⁴. In yeast, some secretory proteins can be secreted in the absence of COPI function, suggesting that there must be another mechanism for their transport⁵. Furthermore, algae secrete large sugar-protein conjugates (called scales), which are processed in the Golgi and can be 20 times the size of a COPI vesicle⁶. These observations argue in favour of the Golgi cisternae carrying the material forward, as proposed in the cisternal maturation model.

But what then do the COPI vesicles do? One proposal is that they carry Golgi proteins in the retrograde direction, recycling resident Golgi proteins from cisternae that are fragmenting at the *trans* face and incorporating them into new layers at the *cis* face. In this scheme, the cisternae break up at the *trans* face to make COPI vesicles, as well as the secretory vesicles that carry secretory proteins for the final step of their journey. The proteins end up

either at the plasma membrane, where they are expelled from the cell, or at an organelle called an endosome. This version of the cisternal maturation model received a boost when Bonfanti *et al.*⁷ showed that 300-nm procollagen bundles, which are eventually secreted, travel forward through mammalian Golgi stacks without leaving the cisternae.

How do proponents of the vesicle-shuttle model interpret these results? Orci and colleagues⁸ proposed that bulky secretory cargoes are transported by ‘mega-vesicles’ that can be substantially larger than conventional COPI vesicles. This is not impossible. The kinetics of membrane fission (the mechanism by which a budding vesicle separates from the membrane on which it forms) might regulate the size of a vesicle, so that bulkier cargoes delay the final fission event until a vesicle of suitable size has been generated⁹. Another argument against cisternal maturation is that, contrary to its predictions, several studies report that resident Golgi proteins are not concentrated in COPI vesicles⁴.

These unresolved issues meant that the cisternal maturation concept remained unproved. But, just as it looked as if the debate would drag on for ages yet, Losev *et al.* (page 1002)¹ and Matsuura *et al.* (page 1007)² have independently produced striking images that purport to show direct evidence for cisternal maturation.

The authors exploited the fact that in the budding yeast *Saccharomyces cerevisiae*, the Golgi cisternae are not stacked. This organization makes it possible to follow cisternal dynamics by video microscopy of live cells. Resident proteins characteristic of early and late Golgi cisternae were tagged with two differently coloured fluorescent tags to mark the different cisternae (for example, green for early cisternae and red for late ones). Remarkably, the tagged cisternae changed colour over time, showing a consistent progression. For instance, a green early Golgi cisterna became gradually yellow and then red, as it lost the early Golgi marker and acquired the late Golgi marker. (In such imaging studies, the presence of both green and red fluorescence gives a yellow colour.) These findings are consistent with a cisternal maturation model for Golgi trafficking in budding yeast.

The story is not yet complete: a few key predictions of the cisternal maturation model remain to be tested. For instance, secretory cargoes should remain within the cisternae as they mature, while the resident Golgi proteins come and go — the obvious experiment to confirm this would be to fluorescently tag secretory cargoes at the same time as resident Golgi proteins and watch their relative progress.

Another prediction is that secretory cargoes present in an early cisterna should move through the Golgi together. Preliminary evidence is that they do: Losev *et al.*¹ report that two yeast secretory cargoes (α -factor and carboxypeptidase Y) traverse the Golgi at roughly the same rate. Moreover, Mironov *et al.*¹⁰ showed that two mammalian secretory cargoes, procollagen and the vesicular stomatitis virus glycoprotein (VSV-G), travel through the Golgi at approximately the same speed. But VSV-G at high concentrations

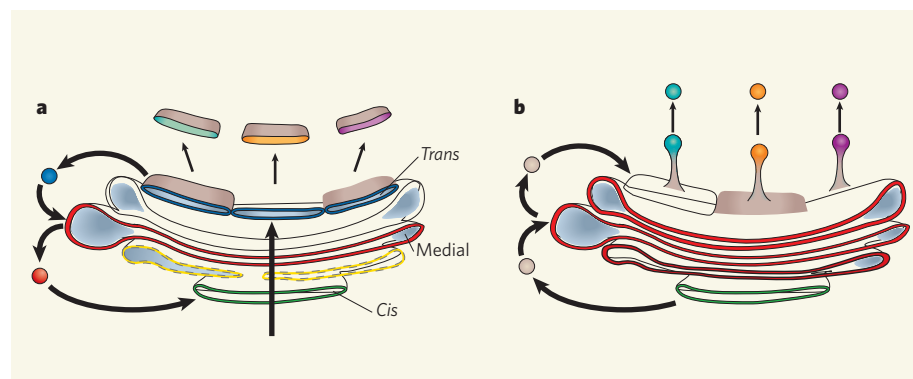


Figure 1 | Two models of protein trafficking through the Golgi. a, In the cisternal maturation mechanism, vesicles derived from the endoplasmic reticulum fuse to form the *cis* cisternae of the Golgi (green). This cisterna then transforms into the medial cisterna (outlined in red) by a maturation process, possibly via an intermediate (yellow). Finally, the medial cisterna matures into the *trans* face of the Golgi (blue), to be consumed during secretory traffic to the cell surface. Sorting proteins for different destinations is accomplished by dividing cargo into different regions of the *trans* compartment of the Golgi before it fragments to generate vesicles for different destinations. COPI vesicles would retrieve and recycle resident Golgi proteins at each step. **b**, In the vesicle-shuttle mechanism, the *cis*, medial and *trans* cisternae of the Golgi remain in place, without maturing, and COPI vesicles deliver cargo between cisternae. At the *trans* compartment, vesicles containing different secretory cargoes bud off to deliver their contents to distinct destinations.

might make large protein complexes, and therefore behave as a large cargo, so additional experiments are needed to look at smaller cargoes in mammalian cells.

The fate of the late cisternae also needs to be clarified. If the cisternal maturation model is correct, they should mature into secretory vesicles and other types of carrier, but this has yet to be confirmed by fluorescence microscopy. A related prediction is that blocking retrograde traffic should block cisternal maturation and send Golgi-resident proteins to the cell surface or the endosome. Surprisingly, Matsuura *et al.*² find that in a yeast mutant with a defect in COPI vesicle assembly, cisternal maturation is slowed about three-fold — but it still occurs. How are resident Golgi proteins being recycled in the absence of COPI function? Do resident Golgi proteins escape from the organelle under these conditions?

Even if cisternal maturation does occur, the details of the mechanism have yet to be defined. For instance, how is the polarized *cis*-to-*trans* distribution of resident Golgi proteins maintained? One suggestion is that it might be by the differential recycling of the Golgi proteins⁴. And what of other cell types? It is quite possible that in *S. cerevisiae*, the Golgi is formed *de novo* and then consumed during each round of secretory protein

transport. In mammalian cells, however, there is no evidence that Golgi membranes form *de novo* under physiologically relevant conditions during sequential rounds of secretion. Maybe these cells rely instead on a variety of trafficking modes commensurate with cargo size and Golgi organization. Whatever the answer may be, the images provided by Losev *et al.*¹ and Matsuura *et al.*² have certainly swung the scales heavily towards cisternal maturation, in yeast at least.

Vivek Malhotra is in the Department of Cell and Developmental Biology, University of California, San Diego, La Jolla, California 92014, USA. Satyajit Mayor is at the National Centre for Biological Sciences, Bangalore 560 065, India. e-mail: vivek@biomail.ucsd.edu

1. Losev, E. *et al.* *Nature* **441**, 1002–1006 (2006).
2. Matsuura, K., Takeuchi, M., Ichihara, A., Mikuriya, K. & Nakano, A. *Nature* **441**, 1007–1010 (2006).
3. Pelham, H. R. B. & Rothman, J. E. *Cell* **102**, 713–719 (2000).
4. Glick, B. S. & Malhotra, V. *Cell* **95**, 883–889 (1998).
5. Gaynor, E. C. & Emr, S. D. *J. Cell Biol.* **136**, 789–802 (1997).
6. Becker, B. & Melkonian, M. *Microbiol. Rev.* **60**, 697–721 (1996).
7. Bonfanti, L. *et al.* *Cell* **95**, 993–1003 (1998).
8. Volchuk, A. *et al.* *Cell* **102**, 335–348 (2000).
9. Diaz Anel, A. M. & Malhotra, V. *J. Cell Biol.* **169**, 83–91 (2005).
10. Mironov, A. A. *et al.* *J. Cell Biol.* **155**, 1225–1238 (2001).

enzyme Rubisco activase^{1,3}, when first CO₂ and then a magnesium ion bind to the active site. The substrate, ribulose biphosphate, then reacts with these to form an enediol intermediate, which engages with either another CO₂ or an O₂ molecule, either of which must diffuse down a solvent channel to reach the active site. Tcherkez and colleagues' achievement² is to have produced an explanatory mechanism for the trade-off usually observed between the specificity factor (that is, a ratio indicating selectivity for CO₂ over O₂, which ranges between 20 and 280; refs 1, 3) and k_{cat} (the rate of enzyme turnover, which varies between two and eight catalytic events per second).

Mutagenesis

Work leading up to this proposed mechanism⁴ involved site-directed mutagenesis in tobacco, which had the aim of destabilizing the enzyme active site and altering the reactivity of the enediol intermediate. Changing the binding of an amino-acid residue that encourages the addition of both CO₂ and O₂ dramatically increased the rate of oxygenase activity.

Such observations^{3,4} provided the key to the idea² that in the active site the enediol must be contorted to allow CO₂ to attack more readily despite the availability of O₂ molecules. The more the enediol mimics the carboxylate end-product, Tcherkez *et al.*² conclude, the more difficult it is for the enzyme to free the intermediate from the active site when the reaction is completed. When the specificity factor and selectivity for CO₂ are high, the impact on associated kinetic properties is greatest: k_{cat} becomes slower.

So, rather than being inefficient, Rubisco has become highly tuned to match substrate availability. Several other correlates are also explained by this relationship. For instance, Rubisco discriminates more against ¹³C than against ¹²C, the two naturally occurring stable isotopes in CO₂. But when the specificity factor is high, the ¹³C reaction intermediate binds more tightly, and so carbon isotope discrimination is higher (that is, less ¹³C is incorporated); in consequence, the carbon-isotope signals used to reconstruct past climates should perhaps now be re-examined. In contrast, higher ambient temperatures (30–40 °C) reduce the stability of the enediol, and Rubisco oxygenase activity and photorespiration rate increase.

These insights² into the mechanism of Rubisco catalysis are timely, because progress is being made in identifying the evolutionary origins of a wider range (forms I to IV) of Rubisco-like proteins⁵. The least effective of these forms have evolved, and now reside, in microorganisms in anaerobic sediments where catalysis does not have to compete with oxygen¹. One bacterium can express all three catalytically active forms (I, II and III), and switches between them depending on environmental conditions⁶. The evolutionary

PLANT BIOLOGY

Designs on Rubisco

Howard Griffiths

Rubisco is said to be both the most important enzyme on Earth and surprisingly inefficient. Yet an understanding of the reaction by which it fixes CO₂ suggests that evolution has made the best of a bad job.

Rubisco is the enzyme in photosynthesis that is responsible for the conversion of inorganic carbon, as CO₂, into organic compounds. The demanding initial catalytic step (or carboxylase reaction¹) precedes the photosynthetic reduction of reaction products using the energy trapped from sunlight. The acronym Rubisco actually stands for ribulose biphosphate carboxylase-oxygenase, because the enzyme also has a tendency to confuse O₂ for CO₂ as its substrate. Rubisco has the reputation of being slow and inefficient, but it is one of life's big successes: globally there is an estimated 5–10 kg Rubisco for every person on Earth, and each year it reacts with 15% of the total pool of atmospheric CO₂.

Active site

Work to 'improve' Rubisco and so increase crop productivity has usually foundered on the catalytic active site, which is highly conserved in different forms of the enzyme and

has generally proved to be intractable to genetic manipulation. But a fresh angle on such prospects comes from Tcherkez and colleagues², writing in the *Proceedings of the National Academy of Sciences*. They propose an explanation for the reaction mechanism that accounts for the selection of CO₂ in preference to O₂. The systematic evolution of enzyme kinetic properties seems to have occurred in Rubisco from different organisms, suggesting that Rubisco is well adapted to substrate availability in contrasting habitats.

It is curious that Rubisco should fix CO₂ at all, as there is 25 times more O₂ than CO₂ in solution at 25 °C, and a 500-fold difference between them in gaseous form. Yet only 25% of reactions are oxygenase events at this temperature, and carbon intermediates 'lost' to the carbon fixation reactions by oxygenase action are metabolized and partly recovered by the so-called photorespiratory pathway. Catalysis begins with activation of Rubisco by the

trade-offs observed by Tcherkez *et al.* are consistent with the occurrence of most of these enzymes. Alternatively, some higher plants and photosynthetic microorganisms have developed mechanisms to suppress oxygenase activity: CO₂-concentrating mechanisms are induced either biophysically⁷ or biochemically^{8,9}. There are systematic changes in Rubisco kinetics^{1,3}, consistent with Tcherkez and colleagues' views, that are thought to have evolved in the past 10 million years in plants related to sugar cane and maize.

Molecular techniques such as 'directed evolution' offer the possibility of manipulating part of Rubisco — the large subunit component — to select for catalytic improvements¹⁰, allowing access to 'sequence space' that is not available to conventional molecular manipulations. A sensible move would be to screen the kinetics of Rubisco enzymes from different plants that inhabit extreme environments^{1,3}. For instance, Rubisco has not been characterized in the so-called CAM plants, which use a form of photosynthesis (crassulacean acid metabolism) adapted for arid conditions. In these species, the enzyme is exposed daily to a range of CO₂ concentrations (0.01–2%) equivalent to those occurring throughout much of Earth's history⁹.

Other research avenues include manipulating the various components of Rubisco¹ and cell-specific targeting of chimaeric Rubiscos⁸. Potential pitfalls here are that the modified Rubisco would not only have to be incorporated and assembled by crop plants, but any improved performance would have to be retained by the plants. Finally, one suggestion is that we should engineer plants that can express two types of Rubisco — each with kinetic properties to take advantage of the degree of shading within a crop canopy¹¹. Such rational design¹⁰ would not only offer practical opportunities for the future, but also finally give the lie to the idea that Rubisco is intransigent and inefficient.

Howard Griffiths is in the Physiological Ecology Group, Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK.
e-mail: howard.griffiths@plantsci.cam.ac.uk

MATERIALS SCIENCE

Relaxors go critical

R. E. Cohen

Relaxor ferroelectrics are fascinating and useful materials, but they seem to be heterogeneous, hopeless messes. Observing what they do under electric fields reveals critical behaviour that helps to make sense of them.

Piezoelectrics — materials that convert electrical energy into mechanical energy, and vice versa — form the alarms in our watches, the warning buzzers in our cars, and the transducers used in sonar and medical ultrasound. They are also being used for knifeless surgery and to build tiny pumps and motors for medical applications. On page 956 of this issue, Kutnjak and colleagues¹ report investigations of the thermodynamic properties and phase diagram of relaxor ferroelectrics with giant piezoelectric effects. These materials offer a response that is up to ten times greater than that of standard piezoelectrics, and they are revolutionizing acoustic imaging and surgical applications.

Ferroelectrics have a spontaneous polarization (a dipole moment per unit volume, or net charge flow per unit volume from a non-polar state) that can be switched in direction by applying an electric field. A prototypical ferroelectric is lead titanate (PbTiO₃, or PT). Ferroelectric behaviour arises as a result of competition between long-range forces between ionic charges in the material, which act to destabilize the nonpolar structure, and short-range, repulsive forces, which have a stabilizing influence. Covalency softens this repulsion and allows the atoms to have average off-centre displacements², and a net polarization. The piezoelectric behaviour of ferroelectrics arises from the coupling of strain and polarization when the polarization interacts with applied electric fields.

Relaxor ferroelectrics are solid solutions between a relaxor material and a ferroelectric such as PT. Relaxors do not have a polar ground state, and are heterogeneous³. They have disordered polarization, with small ordered 'polar nanoregions'^{4,5} that individually polarize⁶ below a temperature known as the Burns temperature⁷.

Kutnjak and colleagues study the relaxor ferroelectric system PMN-PT (in full, PbMg_{1/3}Nb_{2/3}O₃-PbTiO₃), which is an exemplar of a new class of relaxor ferroelectric with pronounced piezoelectric properties. In the normal, collinear piezoelectric case, electric field is applied parallel to the polarization, and the resulting strain is small. But in high-coupling relaxor ferroelectrics, the energy barrier for polarization rotation is low, so even a field applied obliquely can easily rotate the polarization (Fig. 1). It is the polarization rotation effect that gives rise to very large electromechanical coupling^{8,9}, with strains of up to 2% observed. Polarization rotations have been

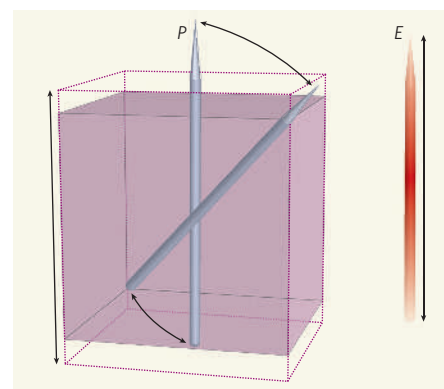


Figure 1 | The polarization rotation effect. When an electric field *E* is applied along the cube axis (vertical), the polarization *P* rotates from the cube diagonal towards the cube axis. With the polarization along the cube diagonal, the lattice strain is small (the lattice is close to cubic), but with the polarization along the cube axis the strain is large — there is a large piezoelectric effect. Kutnjak *et al.*¹ apply a field along the cube diagonal in a new-generation relaxor ferroelectric, and find that a first-order jump in polarization with temperature for small fields vanishes at a critical line that depends on temperature, electric field and the material's composition. The electromechanical coupling also peaks at this critical line.

proved experimentally from X-ray diffraction and optical studies^{10,11}.

Relaxor ferroelectrics are enormously useful because of their piezoelectric properties. But they also exhibit a cornucopia of mesoscopic and microscopic heterogeneities over a range of lengths and timescales. This diversity has so far made it difficult to systematize observations of them or even to determine their whole phase diagram — the jumping-off point for any materials study.

Kutnjak *et al.*¹ measured properties of PMN-PT as functions of temperature, electric field and composition. In the resulting phase diagram, they found a critical line under applied electric field. For low fields, there is a first-order transition where local dipoles formed by atomic displacements are ordered at low temperatures, and rotate or become disordered, as in a paraelectric material, for temperatures above the transition point. At this transition, there is a discontinuity in the net polarization. As electric field is increased, the discontinuity in polarization decreases until it is zero at the critical line. This finding complements recent evidence for critical behaviour in

1. Spreitzer, R. J. & Salvucci, M. E. *Annu. Rev. Plant Biol.* **53**, 449–475 (2002).
2. Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. *Proc. Natl Acad. Sci. USA* **103**, 7246–7251 (2006).
3. Parry, M. A. J., Andralojic, P. J., Mitchell, R. A. M., Madgwick, P. J. & Keys, A. J. *J. Exp. Bot.* **54**, 1321–1333 (2003).
4. Whitney, S. M., von Caemmerer, S., Hudson, G. S. & Andrews, T. J. *Plant Physiol.* **121**, 579–588 (1999).
5. Ashida, H., Danchin, A. & Yokota, A. *Res. Microbiol.* **156**, 611–618 (2005).
6. Yoshizawa, Y., Toyoda, K., Arai, H., Ishii, M. & Igarashi, M. *J. Bacteriol.* **156**, 5685–5691 (2004).
7. Giordano, M., Beardall, J. & Raven, J. A. *Annu. Rev. Plant Biol.* **56**, 99–131 (2005).
8. Brown, N. J., Parsley, K. & Hibberd, J. M. *Trends Plant Sci.* **10**, 215–221 (2005).
9. Dodd, A. N., Borland, A. M., Haslam, R. P., Griffiths, H. & Maxwell, C. *J. Exp. Bot.* **53**, 565–580 (2002).
10. Parikh, M. H., Greene, D. N., Woods, K. K. & Matsumura, I. *Protein Engng Design Select.* **19**, 113–119 (2006).
11. Long, S. P., Zhu, X.-G., Naidu, S. & Ort, D. R. *Plant Cell Environ.* **29**, 315–330 (2006).

PMN–PT at high pressures¹². Unusual properties will emanate away from these bounding critical lines, giving rise to many of the strange and difficult-to-understand properties of these complex materials.

This critical behaviour can be better understood by analogy with the familiar pressure–temperature phase diagram of H₂O. Heating water at 100 °C (373 K) under atmospheric pressure converts it to steam at a phase transition, and the jump in density between steam and water defines the transition as first order. As pressure is increased, the density difference between water and steam decreases, and goes to zero at a pressure of 22 megapascals (and a temperature of 647 K); above this critical point, there is no phase transition. Near the critical point, H₂O has unusual properties, such as large density fluctuations and opalescence. Water would seem a strange substance if we lived at pressures near the critical point (it is anyway, but not because of this particular critical point).

To understand the importance of the critical lines in PMN–PT, even away from the actual critical values of applied field, one must understand that, whereas a fluid such as water cannot sustain large pressure gradients (it will simply flow to equalize the pressure), ferroelectrics can, by their nature as dielectrics, sustain large electric fields. These fields can be generated locally by chemical heterogeneities, domains and surfaces. Thus, even without an externally applied electric field, portions of the sample may be in the critical regime.

Near the critical line described by Kutnjak and colleagues, rotating the polarization becomes extremely easy, and in fact diverges at the critical point. This discovery helps to explain why PMN–PT has such a high coupling strength, and is likely to be a ubiquitous feature of these materials. It used to be thought that the complex chemistry of relaxor ferroelectrics was related to their large coupling; but the prediction that pure PT under pressure would show polarization rotation and a piezoelectric constant larger even than that of the high-strain piezoelectrics suggests otherwise¹³. Delineating these fundamental issues will make it easier to develop new materials and improve existing ones, as well as help to explain multitudes of experimental data.

This is an exciting and collegial field, in which theory, experiment and materials development work side by side. First-principles theory is being used for ‘materials by design’^{14,15}. Basic physical ideas such as the nature of polarization^{16,17}, molecular dynamic simulations⁴ and advanced experimentation^{3,5,10–12} have brought us to a point at which we not only have a greatly improved understanding of basic physical phenomena, but are also poised on the edge of hugely expanded applications of acoustic technologies. ■

R. E. Cohen is at the Carnegie Institution of Washington, 5251 Broad Branch Road, NW, Washington DC 20015, USA.

e-mail: cohen@gl.ciw.edu

- Kutnjak, Z., Petzelt, J. & Blinc, R. *Nature* **441**, 956–959 (2006).
- Cohen, R. E. *Nature* **358**, 136–138 (1992).
- Ye, Z. G. *et al. Phys. Rev. B* **67**, 104104 (2003).
- Burton, B. P., Cockayne, E. & Waghmare, U. V. *Phys. Rev. B* **72**, 064113 (2005).
- Jeong, I. K. *et al. Phys. Rev. Lett.* **94**, 147602 (2005).
- Vakhrushev, S. B. & Shapiro, S. M. *Phys. Rev. B* **66**, 214101 (2002).
- Burns, G. & Dacol, F. H. *Phys. Rev. B* **28**, 2527–2530 (1983).
- Fu, H. & Cohen, R. E. *Nature* **403**, 281–283 (2000).
- Garcia, A. & Vanderbilt, D. *Appl. Phys. Lett.* **72**, 2981–2983 (1998).
- Noheda, B. *et al. Phys. Rev. Lett.* **86**, 3891–3894 (2001).
- Chien, R. R., Schmidt, V. H., Tu, C.-S., Hung, L. W. & Luo, H. *Phys. Rev. B* **69**, 172101 (2004).
- Samara, G. A. & Venturini, E. L. *Phase Transitions* **79**, 21–40 (2006).
- Wu, Z. & Cohen, R. E. *Phys. Rev. Lett.* **95**, 037601 (2005).
- Nakhmanson, S. M., Rabe, K. M. & Vanderbilt, D. *Phys. Rev. B* **73**, 1–4 (2006).
- Dawber, M. *et al. Phys. Rev. Lett.* **95**, 1–4 (2005).
- King-Smith, R. D. & Vanderbilt, D. *Phys. Rev. B* **47**, 1651–1654 (1993).
- Resta, R. *Rev. Mod. Phys.* **66**, 899–915 (1994).

IMMUNOLOGY

A second chance for the thymus

Hans-Reimer Rodewald

Each organ develops at its own time — usually in the embryo. The discovery of progenitor cells that give rise to two structures in the thymus hints that this immune organ can continue to develop after birth.

The thymus is the source of the mature T cells that fight many types of infection. In this issue, two papers^{1,2} identify the potential progenitors that enable its development. Rossi *et al.* (page 988)¹ demonstrate that single embryonic epithelial cells contribute to two major structural components of the mature thymus (the medullary and cortical epithelia). An accompanying paper by Bleul *et al.* (page 992)² shows that thymus epithelial progenitors continue to be active after birth, and probably generate new thymus epithelium constantly. Remarkably, genetic activation of thymus epithelial progenitors in ‘nude’ mice that do not make a functional thymus, results in the *de novo* formation of units of functional thymus.

The immature progenitor cells that give rise to T cells are constantly supplied to the thymus from the bone marrow. On entering the thymus, they divide rapidly before undergoing a series of well-defined developmental steps, requiring tight contact with the thymus structure. Specialized thymus epithelial cells (TECs)^{3,4} nurture the immature T cells by providing growth factors and cell–cell contacts. Moreover, contact of immature T cells with TECs is vital for the selection procedures that ensure that the mature T cells do not react against the tissues of the body itself (which would cause autoimmune disease). TECs come in two major types — medullary TECs (mTECs) and cortical TECs (cTECs) — that differ in their anatomical location in the thymus. During their development, T cells migrate through these distinct regions, and the different TECs have separate functions in T-cell growth, selection and export^{5,6}.

During embryonic development, the thymus arises from the third pharyngeal pouch, which is part of the primitive gut⁷. Classical histology suggested that the thymus is derived from two separate embryological

tissues: the endoderm and the ectoderm. It was originally thought that the typical thymus architecture, with medullary and cortical compartments, results from folding an ectodermal layer around an endodermal layer⁸. However, Gordon *et al.*⁹ suggested that endoderm alone may be sufficient for thymus development.

So, what exactly are the thymus-forming cells? Do thymus stem or progenitor cells exist; and if so, how and when do they contribute to the formation of the thymus? In the classical embryological view, thymus-specific stem cells are not necessarily required. Nevertheless, there has been speculation that epithelial stem cells exist that might give rise to both thymic cortex and thymic medulla cells^{10,11}. When transplanted into adult mice, populations of embryonic or fetal thymus epithelial cells harbour all the components required to generate a functional thymus environment *in vivo*^{12–14}. The presence of both cortex and medulla in such reassembled organs^{13,14} (and the common origin of cortex and medulla in certain ‘chimaeric’ mice¹⁵) has been taken as evidence for thymic epithelial progenitors with the dual potential to make both tissues. With the exception of cell clusters of thymic medulla reported previously as being derived from a single cell¹², there have been no assays to analyse the progenitor–product relationship and the developmental potential of single cells.

Taking complementary cellular and genetic routes, Rossi *et al.*¹ and Bleul *et al.*² searched for further evidence of thymus stem or progenitor cells during embryonic and adult life. Rossi *et al.* used single epithelial cells tagged with yellow fluorescent protein (YFP) isolated from embryonic thymus (at day 12) and followed the cells’ development in a non-fluorescent ‘foster thymus’ over several weeks *in vivo* (Fig. 1a). In every case where fluorescent progenitor cells were detected, the single embryonic

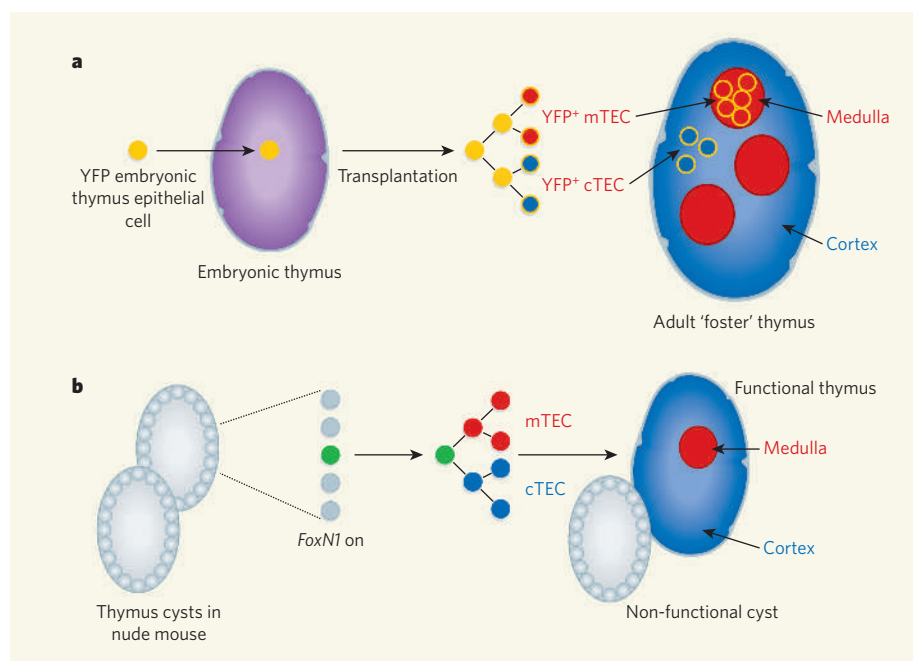


Figure 1 | Thymus development. **a**, Rossi *et al.*¹ tagged single embryonic epithelial cells with yellow fluorescent protein (YFP) and followed their development in the thymus. They found that these progenitor cells can contribute to adult cortical (cTEC) and medullary (mTEC) epithelial cells. Bleul *et al.*² used complex genetics to tag single epithelial cells with YFP postnatally, and found that these older cells can also give rise to both cortical and medullary epithelium. **b**, Mice of the nude strain are unable to make functional thymus because of a defect in the *FoxN1* gene. In these mice, dysfunctional thymic cysts form instead. Bleul *et al.* engineered a novel nude mouse strain in which a normal, functional copy of the *FoxN1* gene was switched on randomly in single postnatal epithelial cell precursors (green). These cells were able to form small 'units' of functional thymus.

epithelial cell had produced both mTECs and cTECs in the adult thymus. This is consistent with epithelial cells in the day-12 thymus containing a large proportion of progenitor cells with this dual potential. One cell can only come from one embryonic tissue. So, these results (supported by those from Bleul *et al.* discussed below) show that both cortex and medulla are derived from a single layer — probably the endoderm⁹.

Bleul *et al.*² used complex genetics to examine the role of epithelial progenitors in postnatal thymus development. The fate of the cells was again followed using a YFP tag, but in this case expression of the fluorescent protein was switched on only after birth. The switch occurred randomly and only in very rare epithelial progenitors, akin to a 'random generator'. The authors' data provide independent evidence that a single epithelial cell can give rise to both types of TEC *in vivo*, but they also show that this dual potential is maintained in the thymus after birth. Some of the thymi showed fluorescent clusters of only medullary or only cortical epithelial cells. This suggests that, in addition to progenitors with dual potential, there are distinct progenitor cells that are committed to either an mTEC or a cTEC fate. The distance of cortical from medullary epithelial cell clusters indicates that these distinct mTEC or cTEC progenitors migrate away from each other after branching off from the common TEC progenitor^{1,2,12,15}.

In a second set of experiments, Bleul and

colleagues attempted the more challenging task of determining whether a single epithelial cell, or very few such cells, can generate a functional thymus. They used the well-known 'nude' mouse¹⁶ which carries a non-functional copy (*FoxN1*^{nu}) of the thymus- and skin-specific gene *FoxN1*. Nude mice have an epithelial cell defect that renders TECs unable to initiate thymus development^{16–18}. As a result, the primitive embryonic thymus develops into cysts composed of a wall of immature TECs and lacking any developing T cells.

Bleul *et al.* constructed a novel nude mouse strain in which the *FoxN1* gene was switched on randomly in single postnatal epithelial cell precursors (Fig. 1b). Remarkably, thymus tissue developed in these mice. This implies that single epithelial progenitors are not only able to contribute to some medullary and cortical epithelial cells but to a fully functional cohort of such cells. The authors have gone a long way to show that these 'neo thymi' can support the development of T cells bearing a diverse T-cell repertoire, and sufficient in numbers to mount a T-cell-dependent immune response.

Several points are worth considering in the light of these papers. The anatomical position of the non-functional thymus cysts in the chest is far away from the original location of the primitive thymus in the pharyngeal pouch, implying that thymus epithelial progenitors can initiate thymus formation outside their normal embryological context. This is remarkable because many reports have stressed

requirements for other cell types for proper thymus development³. It is possible that adult thymus development is independent of such cellular interactions, or that cells surrounding the nude thymus cysts can act in lieu of the embryological counterparts.

Because it was sufficient merely to re-express *FoxN1* in adult life to make a thymus, it would seem that *FoxN1* could be the master regulatory factor in thymus development. The data suggest that, in the absence of *FoxN1* expression, thymus epithelial stem cells become dormant without permanently damaging their viability or developmental potential. Finally, TEC progenitors could continually replace TEC populations in the adult thymus, which might allow the TECs to update the antigens that they use in T-cell selection.

The nature of thymus-forming cells will continue to challenge immunologists because it is likely that thymus epithelial stem cells are not only the cells underlying the formation of the well-known thymus, which is found next to the heart, but are also the origin of the recently noted thymus in the neck^{19,20}. Rossi *et al.* identified most, if not all, embryonic day-12 thymus epithelial cells as progenitors, and Bleul *et al.* demonstrated stem or progenitor activity in the adult thymus. However, isolation and better characterization of these cells is required, and it would be interesting to learn whether self-renewing epithelial stem cells or TEC-committed progenitor cells are responsible for the observed thymus development. ■

Hans-Reimer Rodewald is in the Department of Immunology, University of Ulm, D-89070 Ulm, Germany.

e-mail: hans-reimer.rodewald@uni-ulm.de

- Rossi, S. W., Jenkinson, W. E., Anderson, G. & Jenkinson, E. J. *Nature* **441**, 988–991 (2006).
- Bleul, C. C. *et al.* *Nature* **441**, 992–996 (2006).
- Anderson, G. & Jenkinson, E. J. *Nature Rev. Immunol.* **1**, 31–40 (2001).
- Gray, D. H. *et al.* *Curr. Opin. Immunol.* **17**, 137–143 (2005).
- Petrie, H. T. *Nature Rev. Immunol.* **3**, 859–866 (2003).
- Takahama, Y. *Nature Rev. Immunol.* **6**, 127–135 (2006).
- Manley, N. R. *Semin. Immunol.* **12**, 421–428 (2000).
- Cordier, A. C. & Haumont, S. M. *Am. J. Anat.* **157**, 227–263 (1980).
- Gordon, J. *et al.* *Nature Immunol.* **5**, 546–553 (2004).
- Ritter, M. A. & Boyd, R. L. *Immunol. Today* **14**, 462–469 (1993).
- Blackburn, C. C. *et al.* *Trends Immunol.* **23**, 391–395 (2002).
- Rodewald, H. R., Paul, S., Haller, C., Bluethmann, H. & Blum, C. *Nature* **414**, 763–768 (2001).
- Bennett, A. R. *et al.* *Immunity* **16**, 803–814 (2002).
- Gill, J., Malin, M., Hollander, G. A. & Boyd, R. *Nature Immunol.* **3**, 635–642 (2002).
- Rodewald, H. R. in *Handbook of Stem Cells* (ed. Lanza, R.) 555–564 (Academic, Amsterdam, 2004).
- Nehls, M. *et al.* *Science* **272**, 886–889 (1996).
- Blackburn, C. C. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 5742–5746 (1996).
- Muller, S. M. *et al.* *Proc. Natl Acad. Sci. USA* **102**, 10587–10592 (2005).
- Terszowski, G. *et al.* *Science* **312**, 284–287 (2006).
- Dooley, J., Erickson, M., Gillard, G. O. & Farr, A. G. *J. Immunol.* **176**, 6484–6490 (2006).

BRIEF COMMUNICATIONS

Self-fertilization strategy in an orchid

An orchid that flowers in harsh conditions pollinates itself unassisted by any of the usual agents.

Mating in flowering plants normally relies on animals, wind, gravity or secretion to convey pollen grains from the male (anther) to the female (stigma) organ^{1,2}. Here we describe a new type of self-pollination mechanism in the tree-living orchid *Holcoglossum amesianum*, in which the bisexual flower turns its anther against gravity through 360° in order to insert pollen into its own stigma cavity — without the aid of any pollinating agent or medium. This mode of self-pollination, which occurs under windless, drought conditions when insects are scarce, adds to the variety of mechanisms that have evolved in angiosperms^{3,4} to ensure their reproductive success^{1,5}.

H. amesianum (Orchidaceae) grows on tree trunks at altitudes of 1,200–2,000 m. We monitored the pollination behaviour of the flowers in forests in Simao, Yunnan, China, where flowering occurs from February to April during the drought season, when there is no wind and the relative humidity was $31.5 \pm 3.0\%$ (average of 60 measurements). (For methods, see supplementary information.)

We found that the 1,911 flowers of *H. amesianum* that we observed over three flowering seasons in ten different populations all used the same pollination strategy. An extended rostellum separates the pollen-producing anther and the receptive stigma⁶ in the flower (Fig. 1a, b). When the flower is fully open, the anther cap opens and falls off the column (Fig. 1c), uncovering two pollinia attached to a flexible stipe on the clinandrium. The stipe then rises up (Fig. 1d) and curves forwards and downwards, taking the pollinia across the edge of the rostellum (Fig. 1e); it then curves back and upwards, pushing the pollinia towards the stigma cavity whose receptive surface is facing downwards above the rostellum (Fig. 1f), and finally inserts the pollinia into the stigma cavity (Fig. 1g). The stipe's ring structure helps to secure the pollinia in the stigma cavity to ensure fertilization. Self-pollination may be unsuccessful if the anther cap sticks to the pollinia or if the stipe folds over and prevents the pollinia from reaching the stigma.

Nearly all of the self-pollinated flowers

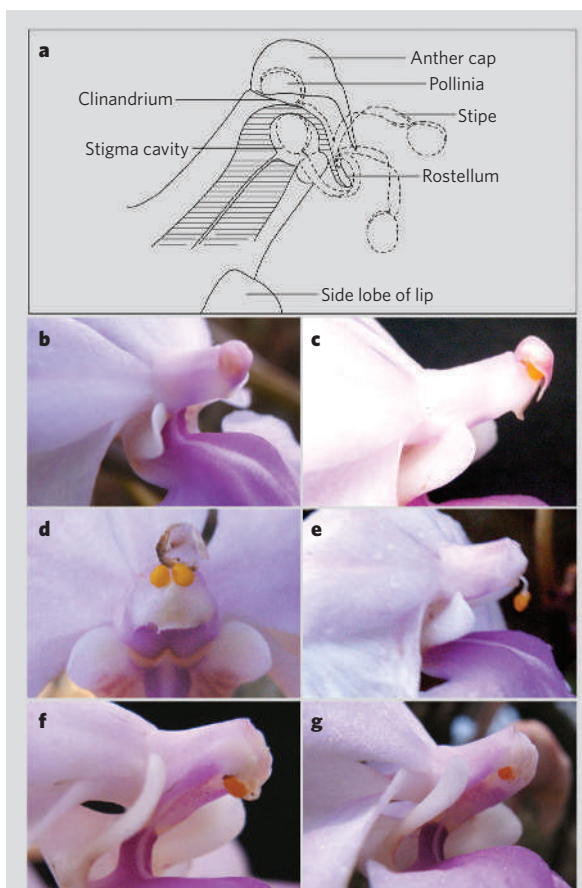


Figure 1 | Mechanism of self-pollination in the orchid *Holcoglossum amesianum*. **a**, Arrangement of floral components and their movements (dashed lines) during the transport of pollen from the anther to the stigma (described in **c–g**). **b**, Open flower before pollination starts; **c**, the anther cap opens and falls off its column; **d**, the stipe carries two pollinia on its tip and rises up from the clinandrium, then, **e**, curves downwards to cross the rostellum; **f**, the stipe next curves up towards the stigma and, **g**, inserts the pollinia into the stigma cavity.

fructified: the rate of fruit set from natural pollination was $50.13 \pm 4.34\%$ ($n = 48$), which corresponds to the success rate for pollinia reaching the stigma. Flowers that failed to complete pollination did not fruit: the non-fruiting rate (49.87%) was equivalent to the total rate of failed anther movement (49.91%). This concordance indicates that intra-flower self-pollination accounts for all fruiting, without cross-pollination or inter-flower selfing⁷.

Fruit-set rates were not significantly different ($F = 0.602$, d.f. = 16, $P > 0.05$) between manual self-pollination ($91.24 \pm 0.88\%$; $n = 9$) and manual cross-pollination ($90.74 \pm 0.83\%$;

$n = 9$), indicating that there were no unfavourable effects of selfing on seed set; or ($F = 2.783$, d.f. = 34, $P > 0.05$) between flowers enclosed in plastic bags before opening until completion of fertilization ($49.49 \pm 3.74\%$; $n = 18$) and unbagged flowers ($51.48 \pm 5.53\%$; $n = 18$), suggesting an absence of outcrossing or inter-flower selfing.

This species is incapable of asexual seed set, as shown by the zero seed set from both bagged and unbagged flowers that had had their anthers removed ($n = 22$); this result for unbagged flowers also shows that no outcrossing or inter-flower selfing was occurring. Apomixis — a type of asexual reproduction involving no fusion of gametes⁸ — is excluded by the presence of antipodal cells in embryos and by the diploid karyotype (two copies of each chromosome), both hallmarks of sexual reproduction, of parents and progeny arising from self-fertilization.

This novel mating mechanism relies only on proactive floral movements against gravity (Fig. 1) — over a period of more than 60 days, we never witnessed any insect visiting the flowers or wind assisting anther movement. It differs from, for example, intra-flower selfing in *Ophrys*², delayed selfing in *Mimulus*⁹, and selfing in the closed flower of *Callitriche*¹⁰ in that it requires no pollinating medium or agent, occurs in open flowers of a land species, and is not a supplemental or back-up mode of pollination. *H. amesianum* has floral features that promote outcrossing such as the spur and rostellum, suggesting that it used to

cross-pollinate, but its spur has become shallow and does not secrete any nectar or odour. Without pollinators for outcrossing, the necessity of ensuring reproductive success must outweigh the potential adverse effects of inbreeding. The present self-contained pollination mechanism is likely to be an adaptation to the orchid's dry and insect-scarce habitat^{8,11} and may be widespread among species growing in similar environments.

Ke-Wei Liu*, Zhong-Jian Liu‡, Lai-Qiang Huang*, Li-Qiang Li‡, Li-Jun Chen‡, Guang-Da Tang‡§

*The Center for Biotechnology and BioMedicine, Graduate School at Shenzhen, Tsinghua University,

Shenzhen 518055, China

e-mail: huanglq@sz.tsinghua.edu.cn

†School of Life Sciences, University of Science & Technology of China, Hefei 230027, China

‡The National Orchid Conservation Center, Shenzhen 518114, China

§College of Forestry, South China Agricultural University, Guangzhou 510642, China

1. Wang, Y. Q., Zhang, D. X., Renner, S. S. & Chen, Z. Y. *Nature* **431**, 39–40 (2004).
2. Darwin, C. *The Various Contrivances by which Orchids are Fertilised by Insects* 2nd edn (Murray, London, 1890).
3. Stebbins, G. L. *Am. Nat.* **91**, 337–354 (1957).

4. Piper, J. G., Charlesworth, B. & Charlesworth, D. *Nature* **310**, 50–51 (1984).
5. Barrett, S. C. H. *Nature Rev. Genet.* **3**, 274–284 (2002).
6. Seidenfaden, G. *Opera Bot.* **114**, 441–444 (1992).
7. Herlihy, C. R. & Eckert, C. G. *Nature* **416**, 320–323 (2002).
8. Holsinger, K. E. *Proc. Natl Acad. Sci. USA* **97**, 7037–7042 (2000).
9. Dole, J. A. *Am. J. Bot.* **79**, 650–659 (1992).
10. Lord, E. M. *Bot. Rev.* **47**, 421–449 (1981).
11. Goodwillie, C., Kalisz, S. & Eckert, C. G. *Annu. Rev. Ecol. Evol. Syst.* **36**, 47–79 (2005).

Supplementary information accompanies this communication on Nature's website.

Received 30 November 2005; accepted 13 April 2006.

Competing financial interests: declared none.

doi:10.1038/441945a

PHOTONICS

Lasers producing tailored beams

Compact lasers that can produce a range of beam patterns are important for progress in several areas, including the improvement of optical tweezers¹, ultra-high-density optical memory² and microfluidics³. Here we engineer photonic crystals to generate semiconductor lasers that produce a range of beam patterns while maintaining stable single-mode oscillation. Our results could contribute to the realization of compact lasers that are capable of producing diverse beam patterns on demand.

The beam (far-field) pattern emitted by a semiconductor laser is determined by the Fourier transformation of the electromagnetic field distribution in its output plane. It is therefore important to find a method of controlling the electromagnetic field while maintaining stable single-mode oscillation. An array of ver-

tical-cavity surface-emitting lasers has been investigated⁴, but without achieving stable and flexible control of the field distribution. The use of a band-edge effect in two-dimensional photonic crystals^{5–8}, in which the refractive index is changed periodically, is a promising way to get this control. The group velocity of light becomes zero at the band edge, which gives rise to the formation of a large and stable two-dimensional single-cavity mode, and the output beam is emitted in the direction normal to the crystal plane, using the crystal itself as a diffraction grating.

We therefore developed photonic-crystal lasers based on the band-edge effect by engineering lattice points⁷ and/or lattice phases^{9,10} in the crystal structure in order to control the internal field distribution (see supplementary information). Scanning electron micrographs of the fabricated photonic crystals are shown in Fig. 1 (a–f, left panels). The engineered photonic-crystal lasers all display stable single-mode oscillation at room temperature; the maximum output power exceeds 45 mW under continuous-wave conditions.

The surface-emitted beams had a variety of shapes, including single, twin and quadruplet doughnuts, either separated or touching, and circular single-lobed forms (Fig. 1, right panels). Divergence angles of these beams were less than 2°, reflecting the large area (about 50 × 50 μm²) of coherent oscillation. An even larger area of coherent oscillation can be achieved by

adjusting the distance between the photonic crystal and the active layer in the device's structure (see supplementary information).

The mechanisms by which the individual beam types are produced can be explained qualitatively. The beam patterns reflect the field distribution within one unit cell of the crystal lattice. Circular lattice points give rise to a rotationally symmetrical field distribution (see supplementary information). Interference in the far field produces a doughnut-shaped beam (Fig. 1a), which can have either tangential or radial polarization (see supplementary information). Shifts of the crystal lattice by half of the lattice constant produce phase shifts in the field distributions: the resultant far-field interactions are reversed, leading to the formation of various types of doughnut beam (Fig. 1b–e, and see supplementary information).

When the structure of the lattice points is modified, the field distribution within one unit cell is altered: for example, with triangular lattice points, destructive interference in the far field does not occur in the *x*-direction, which reflects the asymmetry around the lattice point in this direction; as a result, a circular single-lobed beam can be obtained (Fig. 1f, and see supplementary information).

Further engineering of the photonic crystal structures described here should allow novel beams to be generated. This will define a new direction for semiconductor lasers and could lead to the realization of compact lasers with on-demand beam characteristics.

Eiji Miyai*, **Kyosuke Sakai***, **Takayuki Okano***, **Wataru Kunishi*†**, **Dai Ohnishi*†**, **Susumu Noda***

*Department of Electronic Science and Engineering, Kyoto University, Kyotodaigaku-Katsura, Nishikyo-ku, Kyoto 615-8510, Japan
e-mail: snoda@kuee.kyoto-u.ac.jp

†Optical Device Research and Development Center, ROHM, 21 Mizosaki-cho, Saiin, Ukyo-ku, Kyoto 615-8585, Japan

1. Ashkin, A. *IEEE J. Sel. Top. Quant. Electron.* **6**, 841–856 (2000).
2. Dorn, R., Quabis, S. & Leuchs, G. *Phys. Rev. Lett.* **91**, 233901 (2003).
3. MacDonald, M. P., Spalding, G. C. & Dholakia, K. *Nature* **426**, 421–424 (2003).
4. Raftery, J. J. Jr, Danner, A. J., Lee, J. C. & Choquette, K. D. *Appl. Phys. Lett.* **86**, 201104 (2005).
5. Imada, M. et al. *Appl. Phys. Lett.* **75**, 316–318 (1999).
6. Meier, M. et al. *Appl. Phys. Lett.* **74**, 7–9 (1999).
7. Noda, S., Yokoyama, M., Imada, M., Chutinan, A. & Mochizuki, M. *Science* **293**, 1123–1125 (2001).
8. Letartre, X., Monat, C., Seassal, C. & Viktorovitch, P. *J. Opt. Soc. Am. B* **22**, 2581–2595 (2005).
9. Witjaksono, G. & Botez, D. *Appl. Phys. Lett.* **78**, 4088–4090 (2001).
10. Miyai, E. & Noda, S. *Appl. Phys. Lett.* **86**, 111113 (2005).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Received 28 February 2006; accepted 18 April 2006.

doi:10.1038/441946a

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see Nature contents.

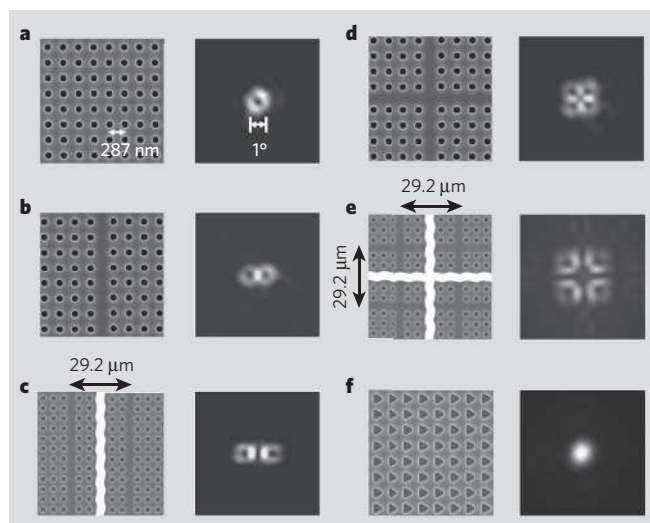


Figure 1 | A range of beam patterns produced by photonic-crystal lasers with engineered lattice points and/or lattice phases. a–f, Scanning electron micrographs of crystal structures (left panels) and observed beam patterns (right panels) for circular lattice points with no lattice shifts, with one shift, with parallel double shifts, with crossed shifts and with double-crossed shifts, and for triangular lattice points with no shifts, respectively. Details, including those of the device structure and of beam-pattern control, are explained in supplementary information.

SLEEP BEHAVIOUR

Sleep in continuously active dolphins

Arising from: O. Lyamin, J. Pryaslova, V. Lance & J. Siegel *Nature* 435, 1177 (2005).

Sleep has been assumed to be necessary for development and to be a vital function in mammals^{1,2} and other animals^{3,4}. However, Lyamin *et al.*⁵ claim that in bottlenose dolphins (*Tursiops truncatus*) and killer whales (*Orcinus orca*), neonates and their mothers show almost no sleep behaviour for the first month after birth; this conclusion is based on their observation that the cetaceans keep swimming, avoid obstacles and rarely close their eyes for 24 hours a day throughout that period. Here we analyse the behaviour and eye closure of three neonate–mother pairs of bottlenose dolphins and find that, although the animals tend to open both eyes when surfacing to breathe, one or both eyes are closed during ‘swim rest’, an underwater sleeping behaviour that is associated with continuous activity. This observation calls into question the conclusions of Lyamin *et al.*⁵, who overlooked this type of sleep by analysing the animals’ eye state only when they surfaced to breathe.

Dolphin sleep is characterized by unihemispheric slow-wave patterns in electroencephalograms, indicating that they may be able to continue swimming even when asleep⁶. Unihemispheric slow waves are invariably linked to closure of the contralateral eye^{7,8} and captive bottlenose dolphins are assumed to be asleep if at least one eye is closed⁹. These animals rest by staying immobile for long periods either on the bottom of the tank (bottom rest) or at the surface (surface rest); most frequently, they rely on ‘swim rest’ — during which they execute a slow swimming motion along a fixed

Table 1 | Closure of right eye by bottlenose dolphins during swim rest and at the surface

		Postpartum age (weeks)			
		1*	4†	5‡	9‡
Mothers	Surface	58.8% (68)	—	42.9% (14)	63.6% (55)
	Underwater	100% (23)	64.5% (62)	72.6% (62)	88.2% (34)
Neonates and babies	Surface	12.1% (99)	—	24.2% (33)	28.5% (63)
	Underwater	91.7% (36)	52.1% (48)	63.3% (60)	88.6% (44)
Adults§	Surface			49.3% (142)	
	Underwater			83.3% (96)	

Percentages indicate the proportion of observational time for which the right eye was closed. Numbers in parentheses give the total number of records. The state of the right eye could be recorded through a viewing window, as in most cases the dolphins swam in an anticlockwise direction round the pool. The percentage of time that the eye was closed at the surface was less than when underwater, especially in neonates; in neonates, it gradually increased with their growth. Observations were supported by Kyeongsoon Kim and Shuhei Hasegawa.

*Female neonate (day 13–14 after birth) at Kamogawa Sea World; born on 9 August 2005.

†Male baby at Minamichita Beachland Aquarium.

‡Male baby at Kamogawa Sea World; born on 20 June 2005.

§Three adult females at Kamogawa Sea World.

trajectory close to the bottom of the tank⁹ (Fig. 1a). We note that Lyamin *et al.*⁵ consider only the immobile states at the surface or on the bottom as typical sleep behaviour.

Two neonate–mother pairs that we monitored for a week or a month after birth took a considerable amount of swim rest (neonates: 31.3–40.7% and 24.6–33.0%; mothers: 32.5–40.7% and 26.7–33.3% of the observational time, respectively). They took this even during the day, when their activity (as estimated from breathing interval and swim speed) was significantly higher than at night⁹. Although nine dolphins, including the neonate–mother pairs, were housed in the same pool, they never

bumped into each other or into the pool wall during swim resting. The occasional emission of click-type sounds recorded during swim rest has indicated that the animals use echolocation to avoid obstacles⁹.

The dolphins that we observed, particularly neonates and babies, tended to close their eyes underwater but to open their eyes when they surfaced to breathe, even during swim rest; the percentage of the observational time during which the eyes were closed was considerably smaller at the surface than when underwater (Table 1). When the 1-week-old neonate surfaced to breathe, it often opened the eye that had been closed underwater (94.7% of 19

M. HIRATA

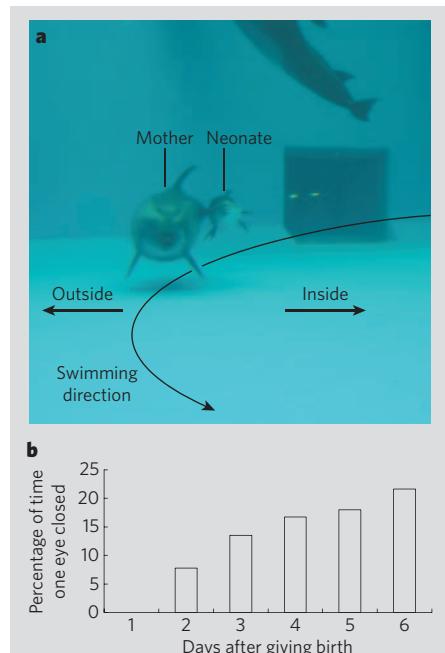


Figure 1 | Trajectory of a neonate–mother pair during ‘swim rest’ and associated eye closure by the mother. **a**, A neonate dolphin and its mother swim-resting at 13 days after birth. **b**, Increase in outer (right-side) eye closure by the mother during the first week after birth (no data were collected on the first day). Observations made with assistance from Satoshi Inoue and Makoto Soichi at Kamogawa Sea World (KSW).

Methods. We studied a female neonate of 2–13 days old and a male baby of 5–9 weeks old, with their mothers, in an oval outdoor pool (18 m by 15 m, 3.5 m deep) that included five other dolphins at KSW in Chiba, Japan; we also studied a 4-week-old male and his mother living in a circular outdoor pool (13 m diameter, 3 m deep) with two other dolphins at Minamichita Beachland Aquarium (MBA) in Aichi, Japan. Their behaviour and eye state were monitored through underwater viewing windows and from the poolside during July–August 2005 in KSW and November 1997 in MBA. The total observational time for each pair was 2,820 min, 780 min and 240 min, respectively. Their behaviour was recorded by point sampling at 30-s intervals for 15 min at the start of each consecutive observation period in KSW and by continuous focal animal sampling for 10 min at the beginning of each hour in MBA. Their eye state in water was recorded by random sampling (5–20 records per hour). Slow ($< 2 \text{ m s}^{-1}$) and circular trajectory swimming near the bottom of the pool was recorded as swim-resting behaviour⁹. To record eye state at the surface and during swim rest, three people with transceivers observed the same animals at the same time from an underwater window (one observer) and from the poolside (two observers at opposite sides of the pool to check both eyes) in KSW. The eye state and position exchange of a neonate–mother pair 2–6 days after birth was examined by focal animal continuous sampling for 2,400 min in KSW. Observations were made between 8:00 and 17:00 to avoid effects of artificial light on the dolphins, particularly neonates. We found previously that a considerable amount of resting and sleeping was evident even during the day⁹.

records, from simultaneous observation of the right eye at the surface and underwater). The neonate opened both eyes at the surface more frequently (69.0% of 29 records) than did the mother (15.8% of 19 records); this may have been because they were not yet fully proficient at breathing.

The first eye closure (unilateral) of the neonate was recorded 14 hours after birth and that of the mother at 11 hours (unilateral) and 108 hours (bilateral) after birth. Unilateral or bilateral eye closure of the mother during swim rest with the neonate was observed for 35.1% of the observation period (1,440 min, 4–6 days after birth). The mother almost always opened the eye facing the neonate (97.8%), indicating that she might be watching the neonate during swim rest (Fig. 1a). The percentage of eye-closure time (right eye) in the mother during swim rest gradually increased after birth from 7.5% (1 day after

birth) to 22% (5 days after birth) (Fig. 1b). Position changes related to unihemispheric sleep^{9,10} were also observed during swim rest in this pair.

Our results show that eye state at the water surface, as monitored by Lyamin *et al.*⁵, is not adequate as an indicator of continuous activity without eye closure: it is likely that the bottlenose dolphins and killer whales described by Lyamin *et al.* also experienced periods of swim-resting sleep. This type of cetacean sleep profoundly differs from sleep behaviour in terrestrial mammals.

Yusuke Sekiguchi*, **Kazutoshi Arai†**,
Shiro Kohshima‡

*Research Center of Physical Fitness, Sports and Health, Toyohashi University of Technology, 1-1 Tempaku Toyohashi-city, Aichi 441-8580, Japan
e-mail: ysekiguc@coepr.tut.ac.jp

†Kamogawa Sea World, 1464 Higashicho Kamogawa-city, Chiba 296-0041, Japan

‡Department of Biological Sciences, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 2-12-1-W3-43 Oookayama Meguro-Ku, Tokyo 152-8551, Japan

1. Hoppenbrouwers, T. & Sterman, M. B. *Exp. Neurol.* **49**, 822–838 (1975).
2. Roffwarg, H. P., Muzio, J. N. & Dement, W. C. *Science* **125**, 604–619 (1966).
3. Shaw, P. J., Tononi, G., Greenspan, R. J. & Robinson, D. F. *Nature* **417**, 287–291 (2002).
4. Rechtschaffen, A., Gilliland, M. A., Bergmann, B. M. & Winter, J. B. *Science* **221**, 182–184 (1983).
5. Lyamin, O., Pryaslova, J., Lance, V. & Siegel, J. *Nature* **435**, 1177 (2005).
6. Mukhametov, L. M. *Exp. Brain Res.* **8**, 227–238 (1984).
7. Lyamin, O. I., Mukhametov, L. M. & Siegel, J. M. *Arch. Ital. Biol.* **142**, 557–568 (2004).
8. Lyamin, O. I. *et al. Behav. Brain Res.* **129**, 125–129 (2002).
9. Sekiguchi, Y. & Kohshima, S. *Physiol. Behav.* **79**, 643–653 (2003).
10. Goley, P. D. *Mar. Mamm. Sci.* **15**, 1054–1064 (1999).

doi:10.1038/nature04898

SLEEP BEHAVIOUR

Activity and sleep in dolphins

Arising from: O. Lyamin, J. Pryaslova, V. Lance & J. Siegel *Nature* **435**, 1177 (2005).

According to Lyamin and co-authors, neonate bottlenose dolphins (*Tursiops truncatus*) almost never sleep, unlike all other mammals that have been studied¹. Although we agree that young dolphins never stop and float at the surface, we find that they spend a considerable amount of time asleep while swimming. Our findings therefore call into question the conclusions of Lyamin *et al.*¹

It is known that bottlenose dolphins use only unilateral sleep and that this state is com-

patible with swimming and with surfacing to breathe^{2–5}; it is also known that young bottlenose dolphins never stop at the surface to rest⁶, as adults sometimes do. However, this should not be interpreted as an indication of sleep absence, particularly as 'swim rest' is the most frequent resting behaviour in adult bottlenose dolphins^{7,8}.

We observed a healthy bottlenose dolphin male calf and its mother during the first year of its life and compared its rest behaviour with that

of a sub-adult male (4 years old) who formed a pair with his 8-year-old brother, and also with that of an adult female (about 20 years old) who formed a pair with her daughter (4 years old). We monitored and measured the incidence of the two main rest behaviours identified in dolphins: slow stereotypic circular swimming and quiescent hanging behaviour^{6–9}.

During slow stereotypic circular swimming, the dolphin moves slowly, following a regular trajectory, and surfaces only to breathe; the animal shows no interest in its environment and produces no sound. Most of the time, one eye is closed while the other performs a sentinel function. Such swim-resting can be displayed by a single individual but it is more usual for two dolphins to pair off and rest together, swimming side by side in close synchrony. Each dolphin keeps one eye open; the open eye is constantly the eye that is directed towards the partner. This pairing behaviour is typical of the mother–calf relationship, but adults also seem to do it in order to save energy and to maintain a state of vigilance while resting.

During quiescent hanging behaviour, the dolphin hangs at the water surface, keeping its blowhole above it. No active swimming is evident, but the animal still has to balance to maintain its position. The animal again shows no interest in its environment and makes no sound. The state of the eyes during this behaviour is hard to monitor and is often uncertain⁶.

We monitored the calf's behaviour using 24-hour underwater video–acoustic recording over 20 separate days during its first year. The sub-adult male was observed for 96 hours during the day and night, and the adult female was observed for 48 hours during the day and night. The state of the calf's eyes was monitored during diurnal hours (8:00–20:00), and those of the sub-adult male and of the adult female were directly observed during diurnal and nocturnal hours (for 96 and 48 hours, respectively).

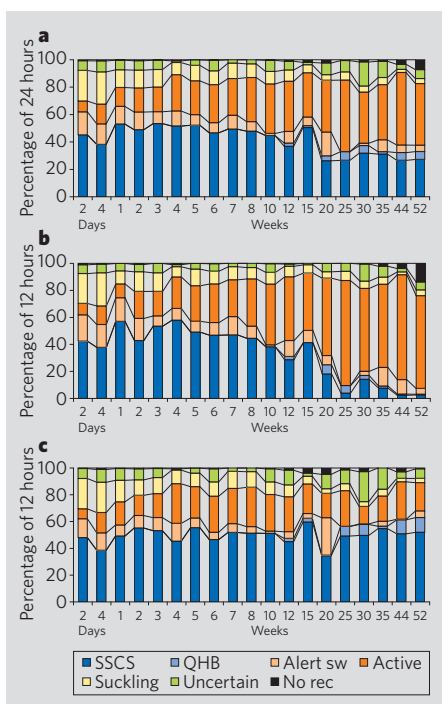


Figure 1 | Trends in the rest behaviour over different periods of a dolphin calf during its first year. **a–c**, Rest trends were monitored **a**, over 24 hours; **b**, during the day-time (8:00–20:00); and **c**, during the night (20:00–8:00). SCS, slow stereotypic circular swimming; QHB, quiescent hanging behaviour; 'alert sw', alert swimming (when the calf is swimming while awake but is not involved in any specific activity); active, playful and exploratory activity; suckling, nursing activity; no rec, behaviour not recorded. **Methods.** Observations were made at the Genoa Aquarium from August 2002 to August 2003. Dolphins were housed in a pool with a large transparent acrylic wall (23.5 m by 8–10 m; 5 m deep; volume, 1,104 m³). The pool was illuminated during the night with three 150 W lamps (9.3 lux at the water surface). Data were collected on 20 days by underwater video–acoustic recording continuously over 24 hours. A single researcher analysed all the video–acoustic sequences. The behavioural categories were measured by using a 1–0 sampling technique during 1-min periods. Further details are available from the authors.

The newborn carried out slow stereotypic circular swimming during the day and night for about 12 hours a day. After one year, the dolphin rested (both hanging and swim resting) for about 8 hours a day, mainly during the night time (Fig. 1). The sub-adult male rested for about 7 hours in 24 and the adult female for about 9 hours in 24; both took their rest mainly during the night. In these three animals, slow stereotypic circular swimming was the most frequent type of rest behaviour observed (constituting at least 85% of the total rest time) and was nearly always associated with unilateral eye closure (90–100%), in agreement with previous observations.

It is generally agreed that dolphins are able to rest and sleep under different and adverse conditions by using different rest strategies. In this respect, the total rest that an individual takes may change, depending on age, health and the physical and social environmental conditions^{6,8}. Nevertheless, the rest behaviour in the dolphin calf that we observed was clear, showing a negative diurnal trend that was only partly compensated for by a nocturnal positive trend. This behaviour resembles that of (diurnal) terrestrial mammals, including humans.

Guido Gnone, Tiziana Moriconi, Giorgia Gambini
Acquario di Genova, Area Porto Antico Ponte

Spinola, 16128 Genova, Italy
e-mail: ggnone@acquariodigenova.it

1. Lyamin, O., Pryaslova, J., Lance, V. & Siegel, J. *Nature* **435**, 1177 (2005).
2. Mukhametov, L. M. *Exp. Brain Res.* **8**, 227–238 (1984).
3. Mukhametov, L. M. *Sleep Res.* **24**, 202 (1995).
4. Mukhametov, L. M., Supin, A. Y. & Polyakova, I. G. *Brain Res.* **134**, 581–584 (1977).
5. Mukhametov, L. M., Oleksenko, A. I. & Polyakova, I. G. *Neurophysiology* (transl. *Neirofiziolgiya*) **20**, 398–403 (1988).
6. Gnone, G., Benoldi, C., Monsignori, B. & Fognani, P. *Aquat. Mamm.* **27**, 29–33 (2001).
7. Flanigan, W. F. *Jr Sleep Res.* **3**, 84 (1974).
8. Sekiguchi, Y. & Kohshima, S. *Physiol. Behav.* **79**, 643–653 (2003).
9. Mukhametov, L. M. & Lyamin, O. I. *J. Sleep Res.* **3**, 174 (1994).

doi:10.1038/nature04899

SLEEP BEHAVIOUR

Lyamin et al. reply

Replying to: Y. Sekiguchi, K. Arai & S. Kohshima *Nature* **441**, doi:10.1038/nature04898 (2006);
G. Gnone, T. Moriconi & G. Gambini *Nature* **441**, doi:10.1038/nature04899 (2006).

All terrestrial mammals studied so far do maximal amounts of sleeping at birth, with sleep time gradually decreasing to adult levels. This has led to the concept that sleep, with its characteristic immobility and unresponsiveness, is necessary for brain and body development. We reported¹ that dolphins and killer whales have a very unusual developmental pattern: neonates are maximally and continuously active at birth, and this activity diminishes over a period of months to the adult level; in the postpartum period, mothers also abruptly cease the characteristic 'hanging' that constitutes the typical sleep behaviour in bottlenose dolphins and killer whales^{2–5}. We did not claim that all sleep was abolished in the postpartum period. Rather, we reported that immobility was absent and that typical sleep posture increases with age, a pattern opposite to that seen in all land mammals studied so far. Sekiguchi *et al.*⁶ and Gnone *et al.*⁷ challenge our conclusions.

Figure 1 of Sekiguchi *et al.*⁶ shows that maternal eye closure at the surface increases with age, which is consistent with our findings. Like us, they find that swimming is well coordinated, that neither neonate nor mother collide with the other animals, that animals vocalize during swimming, and that there is no hanging behaviour in the immediate postnatal period. We have serious reservations about their behavioural and eye-closure data because of the very small number of observations, the making of observations only when the animals approached a single underwater observation window, and the observations of eyelids only during the day rather than during the normal nocturnal sleep period (which also applies to the observations made by Gnone *et al.*⁷). By contrast, we monitored the animals' behaviour continuously and recorded the state

of both eyes during both day and night.

Gnone *et al.*⁷ conclude from their findings that the observed behavioural pattern "resembles [the sleep attitude] of (diurnal) terrestrial mammals, including humans". However, this claim is contradicted not only by our results¹ and by those of Sekiguchi *et al.*⁶, but is also inconsistent with their own data: they report an absence of hanging behaviour in the neonate and mother, with this behaviour not appearing in the neonate until 12 weeks of age and then increasing until one year of age; they also claim that alert swimming is maximal at birth and decreases with age. Both of these observations⁷ agree with our earlier findings¹.

This claim of Gnone *et al.*⁷ also ignores research documenting dolphins' unihemispheric slow waves, avoidance of obstacles while in motion for 24 hours a day, and likely absence of sleep characterized by rapid eye movement (REM)⁸. However, we completely agree with the conclusion of Sekiguchi *et al.*⁶ that cetacean sleep differs profoundly from that of terrestrial mammals.

One might try to draw an analogy between the neonates' continuous activity and sleep walking. However, sleep walking is a pathological, episodic state, in which coordination is grossly impaired and injuries frequently occur; it has not been established that sleep walking effectively substitutes for normal sleep, and it is interspersed with normal sleep episodes. In contrast, the postpartum swimming of dolphins and killer whales is characterized by precise course adjustments and vocalizations as the animals circumnavigate the irregular dimensions of the pools¹. Neonates surface to breathe at 10–40-second intervals, keeping both eyes open, as also described by Sekiguchi *et al.*⁶; such open eyes are an indicator of electroencephalogram-defined wakefulness⁵.

After each breath, the neonate locates, pursues and catches up with the mother, who breathes less frequently. It seems reasonable to us to state that typical sleep behaviour as seen in adult cetaceans and land mammals^{4,5} is greatly reduced in dolphins and killer whales in the postpartum period⁸, and that sustained periods of sleep lasting 10–40 seconds do not occur. Such interrupted sleep does not fulfil sleep's function in land mammals^{1,8}.

Whatever terminology is used, it is clear that the sensorimotor coordination seen in postnatal dolphins and killer whales not only involves extensive muscular activation, but also necessitates widespread activation of the brainstem^{9–11} and most likely of forebrain neurons — in contrast with the neuronal activity pattern seen in the sleep of terrestrial mammals¹². This difference in postpartum behaviour and neural function between cetaceans and terrestrial mammals has important implications for theories of sleep function and particularly for theories of the role of sleep in development⁸.

O. I. Lyamin*†, J. Pryaslova†, V. Lance‡, J. M. Siegel*

*University of California, Los Angeles & Veterans Administration Greater Los Angeles Healthcare System, Sepulveda, California 91343, USA
e-mail: jsiegel@ucla.edu

†Utrish Dolphinarium, Moscow 119071, Russia

‡San Diego State University, San Diego, California 92182, USA

1. Lyamin, O., Pryaslova, J., Lance, V. & Siegel, J. *Nature* **435**, 1177 (2005).
2. Flanigan, W. F. *Sleep Res.* **3**, 84 (1974).
3. Flanigan, W. F. *Sleep Res.* **4**, 139 (1975).
4. Lyamin, O. I. *et al. Behav. Brain Res.* **129**, 125–129 (2002).
5. Lyamin, O. I., Mukhametov, L. M. & Siegel, J. M. *Arch. Ital. Biol.* **142**, 557–568 (2004).
6. Sekiguchi, Y., Arai, K. & Kohshima, S. *Nature* **441**, doi:10.1038/nature04898 (2006).
7. Gnone, G., Moriconi, T. & Gambini, G. *Nature* **441**, doi:10.1038/nature04899 (2006).
8. Siegel, J. M. *Nature* **437**, 1264–1271 (2005).
9. Siegel, J. M. & McGinty, D. J. *Science* **196**, 678–680 (1977).
10. Siegel, J. M. *Brain Res. Rev.* **1**, 69–105 (1979).
11. Siegel, J. M. & Tomaszewski, K. S. *J. Neurophysiol.* **50**, 696–716 (1983).
12. Siegel, J. M. in *Principles and Practice of Sleep Medicine* (eds Kryger, M. H., Roth, T. & Dement, W. C.) 120–135 (Elsevier Saunders, Philadelphia, 2005).

doi:10.1038/nature04900

Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*

Thomas Mitchell-Olds¹ & Johanna Schmitt²

Genomic studies of natural variation in model organisms provide a bridge between molecular analyses of gene function and evolutionary investigations of adaptation and natural selection. In the model plant species *Arabidopsis thaliana*, recent studies of natural variation have led to the identification of genes underlying ecologically important complex traits, and provided new insights about the processes of genome evolution, geographic population structure, and the selective mechanisms shaping complex trait variation in natural populations. These advances illustrate the potential for a new synthesis to elucidate mechanisms for the adaptive evolution of complex traits from nucleotide sequences to real-world environments.

The beginning of the twenty-first century is an exciting time for biologists. Rapid advances in genomics have changed our view of the biological world and fostered new links between molecular biology, ecology and evolution. Genomic studies of natural variation in model organisms are a crucial ingredient in this new synthesis. Molecular biologists have begun to exploit natural variation to identify the genetic mechanisms underlying complex traits¹. Simultaneously, these new genomic tools make it possible for evolutionary biologists to study how ecologically important complex traits evolve in natural environments. These advances now make it possible to understand the adaptive evolution of complex trait variation from molecular mechanisms to geographic patterns of population structure and natural selection.

The diminutive weed *Arabidopsis thaliana* provides an ideal system for such interdisciplinary synthesis. This species—a close relative of *Brassica* crops such as mustard and broccoli—is a convenient genetic model because of its short generation time and small genome. The *A. thaliana* genome was the first plant genome to be sequenced, and the genes and developmental pathways controlling ecologically important traits such as germination, flowering time, pest resistance, and stress tolerance are rapidly being elucidated. It is therefore possible to identify ‘candidate’ genes for adaptive variation in natural populations. *A. thaliana* is a widespread annual weed of rocky places and disturbed sites, native to Europe and central Asia and naturalized in North America (Fig. 1). Across this geographic range, it experiences a broad range of climatic conditions² and selective pressures. Inbred stocks are available for many natural *A. thaliana* accessions (‘ecotypes’), originating across the species’ range. Because the species is habitually inbreeding, genomic and phenotypic data can be combined from multiple experiments with the same genotypes. These genomic tools and resources have enabled a number of important advances in molecular and evolutionary genetics. Here we focus on advances in three complementary areas: (1) genomic studies of molecular variation and population structure; (2) identification of genetic polymorphisms underlying natural variation in complex traits; and, (3) ecological and evolutionary studies of

natural selection and adaptation. Taken together, these advances now make it possible to identify the genetic mechanisms underlying the adaptive evolution of complex traits in natural populations.

Molecular variation and population structure

How much molecular polymorphism exists in *A. thaliana*? Recent genome-wide studies show that an average pair of alleles differs at about seven nucleotides per kilobase (kb; nucleotide diversity = 0.007; refs 3, 4). This is about 50% lower than polymorphism in the outcrossing congener *A. lyrata* ssp. *petraea*⁵, roughly the same as *Drosophila melanogaster*⁶, and nearly an order of magnitude higher than humans⁷. *A. thaliana* has a high frequency of self-pollination in the wild⁸, hence individuals are homozygous at most loci³. Such high rates of self-pollination may influence patterns of linkage disequilibrium, which provides the basis for association studies and linkage disequilibrium mapping in human genetics and plant breeding (see Box 1).

Patterns of nucleotide polymorphism contain a signature of historical demography and natural selection (see Box 2). Genome-wide information for *A. thaliana* has enabled fundamental advances in our understanding of the evolutionary processes that influence these patterns. In many species, a positive correlation exists between local recombination rates and levels of nucleotide diversity³. In regions of low recombination, genetic variation may be reduced owing to ‘hitchhiking’ of neutral variation with nearby selected sites, which will influence wider chromosomal regions when recombination is low. However, this reduction in variation could be due either to selective sweeps of advantageous mutations, or to background selection, which eliminates deleterious mutations and the haplotypes that carry them. Genome-wide polymorphism data make it possible to distinguish these possibilities. The recent observation³ that nucleotide polymorphism is negatively correlated with gene density supports background selection as the predominant mechanism. Gene-dense regions show little sign of rare variants attributable to recent selective sweeps, but the frequencies of non-synonymous polymorphisms indicate purifying selection against deleterious mutations⁹.

¹Department of Biology, PO Box 91000, Duke University, Durham, North Carolina 27708, USA. ²Department of Ecology and Evolutionary Biology, Box G-W, Brown University, Providence, Rhode Island 02912, USA.

Nevertheless, surveys of *A. thaliana* ecotypes have identified loci that may have undergone recent positive selection or selective sweeps (Box 2)¹⁰. Certain other loci show unusually high levels of amino-acid variation or too many intermediate-frequency nucleotide polymorphisms, which exceed neutral expectation and may have been maintained by natural selection^{11–13}. The selective mechanisms maintaining such polymorphisms remain an important open question. Allelic polymorphisms may be maintained within populations by frequency-dependent selection (where common genotypes have low fitness, and rare types are favoured), by temporal variation in selection across seasons and years, or by epistatic selection in which the fitness of an allele depends on genetic background¹⁴. Alternatively, different alleles might be maintained in different populations by local adaptation to geographically divergent selection, resulting in excess polymorphism species-wide, but little variation within individual populations. Wide surveys of *A. thaliana* ecotypes cannot distinguish between these different ecological mechanisms, because information is also needed about patterns of variation within undisturbed natural populations. However, human disturbance and admixture of different source populations may complicate such studies in most parts of the *A. thaliana* range. It is therefore important to understand the geographic population structure of natural variation within the species, as well as the distribution of molecular variation within and among populations across the species range.

Genome-wide polymorphism data for ecotypes across the range of *A. thaliana* have made it possible to investigate the geographic structure of molecular variation. The genetic divergence and geographic distance between pairs of populations is positively correlated across the native range^{3,15}. Such 'isolation by distance' reflects long-term geographical isolation with limited gene flow. The pattern of population ancestry of European ecotypes suggests that much of the native range was colonized from several glacial refugia, with admixture in the zones colonized from more than one source. However, recent human disturbance tends to homogenize variation among populations, especially in agricultural regions of Europe and introduced populations in North America^{15–17}.

How is molecular polymorphism distributed within and among natural populations of *A. thaliana*? Recent studies have estimated the

proportion of total molecular polymorphism that is found within populations to be near 45% in western Europe and North America^{16–18}, but only 12% in less-disturbed parts of Norway¹⁹. Although part of this difference may be attributable to loss of variability as *A. thaliana* migrated north from circum-Mediterranean refugia, comparisons across a 900-km north–south transect in Norway find no latitudinal changes in genetic variation. Therefore, low variation within Norwegian populations is probably not attributable to post-Pleistocene founder events. Instead, high variability within western European populations probably reflects admixture resulting from human disturbance¹⁸. Such admixture complicates our ability to understand the evolutionary forces shaping genetic variation within *A. thaliana* populations.

Finding the genes underlying trait variation

Natural selection operates on complex trait phenotypes. In order to understand the evolution of ecologically important variation, we

Box 1 | Association studies

Association studies use linkage disequilibrium to identify polymorphisms that may be responsible for complex trait variation. Linkage disequilibrium quantifies statistical correlations between polymorphic sites: when linkage disequilibrium is present then information from one locus provides some predictive information about the genotype at a second locus. Association studies take advantage of recombinations accumulated over thousands of generations, and hence may aid identification of individual genes responsible for QTL. In the past (top panel), a new mutation (triangle) occurs at a quantitative trait locus. The original population contains an ancestral chromosome on which the new mutation occurred (black line), as well as other chromosomes (grey lines). After many generations of recombination, the present-day population (lower panel) contains short chromosome regions derived from the original population. Molecular markers near the causal polymorphism (*b*) will be correlated with phenotype, whereas distant markers (*a*, *c*) are uncorrelated because they have been reshuffled by recombination. On average, in *A. thaliana* polymorphisms separated by >50 kb are usually statistically independent, whereas linkage disequilibrium is substantial between sites separated by <50 kb (ref. 71). Because these regions in disequilibrium typically contain in the order of 10 genes, linkage disequilibrium mapping and association studies with candidate genes bring us close to the level of individual loci. These approaches may be useful when pedigrees are small, as in human populations, or when genotyped populations can be used by many researchers for analysis of multiple phenotypes, as in *A. thaliana*. Several challenging statistical issues remain for association studies in *A. thaliana*. Population and pedigree structure can be incorporated into analytical models⁷² or homogeneous populations may be identified. However, if ancestral populations have diverged for the trait of interest, then statistical adjustment for population structure may remove the very effects of interest⁷³. Linkage disequilibrium mapping and association studies also result in false negatives and false positives⁷⁴. Consequently, results from association studies represent hypotheses that must be tested by independent experiments⁴³.

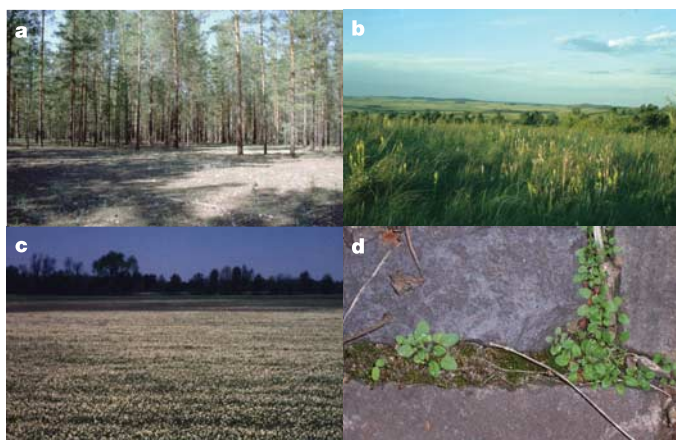
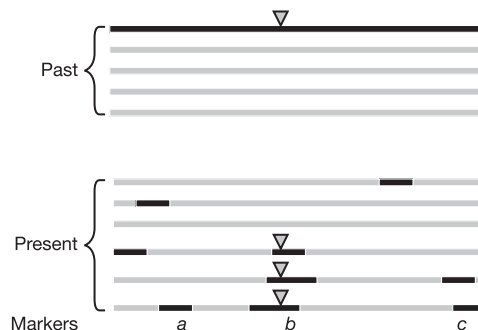


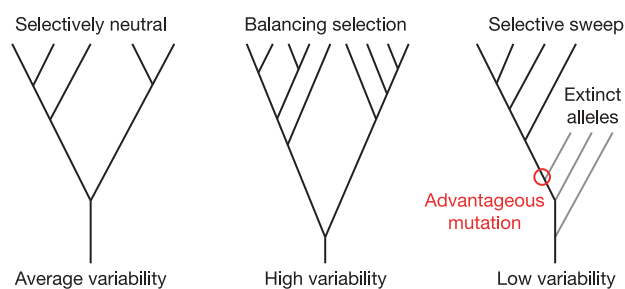
Figure 1 | *Arabidopsis* habitats. **a, b**, Although commonly found in disturbed sites, *A. thaliana* also grows in pine forests on sandy soils (**a**; Bastan, western Siberia) and perennial grasslands in sites with sparse vegetation (**b**; Stepnoje, western Siberia). **c**, Introduced populations in North America may number in the millions. Here, the white flowers are another species, but the interspersed vegetation is predominantly *A. thaliana* (Michigan, USA). **d**, Even in western Europe it can be a long-term resident of stone walls and other sites (Cologne, Germany). (Russia photographs by M. Hoffmann; USA photograph by J. Bergelson; Germany photograph by A. Wilczek.)

must identify nucleotide polymorphisms with functional effects on phenotypic differences. Such polymorphisms are also of great interest to plant molecular biologists seeking to elucidate gene function and genetic pathway structure²⁰. Identification of natural variants is particularly important for functional studies because some knockout mutations may be lethal, or may lack detectable phenotypic effects due to genetic redundancy²¹. These efforts will be greatly aided by databases on genome-wide patterns of nucleotide polymorphism³, gene expression²², and developmental phenotypes²³.

The identification of natural polymorphisms controlling variation in complex traits begins with screens of trait variation among accessions. Once such variation is identified, the next step is to identify polymorphic genomic regions (quantitative trait loci, or QTL) associated with that variation. Recent advances in characterizing genome-wide polymorphism and population structure for a wide range of *A. thaliana* ecotypes^{3,15} may facilitate identification of QTL through linkage disequilibrium mapping (Box 1). Recombinant inbred lines (RILs) between divergent parental ecotypes have already

Box 2 | Sequence signatures

Until recently, cloning of QTL has been confined to laboratory model systems, hence it has been impossible to measure the fitness effects of QTL alleles in natural environments. As an alternative approach, adaptive significance has been inferred from the sequence signature of nucleotide polymorphism⁵. Population genetics theory can predict patterns of nucleotide polymorphism on the basis of standard neutral models for ideal, equilibrium populations. Statistical hypothesis tests can then compare predictions from these standard neutral models to observed variation at genes of interest. For example, the figure shows 'gene trees' portraying the historical relationship among alleles at a given locus. A typical selectively neutral locus is shown on the left, with average levels of nucleotide polymorphism. In contrast, balancing selection (centre) can maintain high levels of variability for long periods of time. Such polymorphisms are unusually old, and have accumulated high levels of molecular variation. Finally, an advantageous mutation may out-compete existing alleles in a 'selective sweep' (right), reducing extant variation below neutral levels. Although these theoretical predictions apply to ideal populations, non-standard or non-equilibrium demographic conditions such as gene flow or changing size⁵ also influence allelic variation in real-world populations. With data from only a single locus, statistical tests often cannot determine whether differences between observed and predicted patterns are due to natural selection, or instead reflect neutral effects of demographic history. Fortunately, genome-wide patterns of variation contain a signature of historical demographic processes, because demography influences variation across the genome. In contrast, the effects of natural selection are confined to the target locus and adjacent regions. Consequently, putative natural selection at genes of interest may be identified by comparison to a genome-wide 'empirical null distribution' summarizing variability at many loci, or by simulations based on plausible non-standard models. Recently, several studies^{3,4} have shown that variation in *A. thaliana* does not conform to standard demographic models of ideal populations. Instead, gene flow among populations, changing population size, extinction, and recolonization evidently have influenced polymorphism in *A. thaliana*.



proved extremely valuable for mapping QTL for a variety of ecologically important complex traits²⁰. QTL of moderately large effect have been identified in a number of studies, but chromosomal regions with large phenotypic effects may actually contain multiple linked QTL of smaller effect²⁴. A substantial portion of genetic variation for complex traits may actually correspond to polygenic models of complex trait variation, with many genes of small effect^{25,26}.

Recent advances in genomics make it possible to treat genome-wide expression patterns as complex traits and screen for natural variation in those patterns. Schmid *et al.*²² studied expression of 22,000 genes in many developmental stages and tissues, providing invaluable data on regulatory patterns in *Arabidopsis* and enabling further functional²⁷ and evolutionary analyses. Patterns of gene expression differ among *Arabidopsis* accessions^{28,29}, and QTL influencing *cis*- or *trans*-regulatory polymorphisms can be identified³⁰. *Cis*-acting allele-specific transcriptional differences are apparently common in *Arabidopsis*^{30,31} and other organisms³². Progress in this field requires experiments that relate transcriptional polymorphisms to phenotypic variation at the whole-plant level.

One important lesson from QTL studies with *A. thaliana* is that the expression of phenotypic differences between QTL alleles may depend strongly on environmental conditions and genetic background. QTL–environment interactions are common; often, certain QTL are expressed in some environments but not in others, or differ in the magnitude of their allelic effects. Although such observations are common in *A. thaliana*^{33,34}, their molecular basis and relationship to genotype–environment interactions at the whole-organism level have rarely been elucidated. To date, antagonistic pleiotropy—in which the alleles of a QTL change phenotypic rank across environments³⁵—seems to be rare in *A. thaliana*. Clearly, more work is needed to understand the genetic basis of genotype–environment interaction.

Epistasis, or interaction effects among different loci, is a fundamental force in many aspects of adaptive evolution³⁶. At its simplest, epistasis occurs when the genotype at one locus influences the magnitude of allelic effects at another locus. More intriguingly, with sign epistasis the genotype at one locus may reverse the direction of allelic effects at another gene. There is now strong evidence that both kinds of epistasis are important in *A. thaliana*^{34,37–40}. In some cases, the mechanisms of epistasis are known from functional studies^{41,42}. At the level of individual genes, sign epistasis influencing juvenile growth rate (an important component of fitness) is caused by genetic polymorphism at a serine/threonine protein kinase, and patterns of nucleotide polymorphism suggest that balancing selection maintains elevated levels of genetic variation²⁵. These observations suggest that epistasis may have a fundamental role in the evolution of *A. thaliana* and other inbreeding species.

Once QTL regions have been identified, the next step is to find the underlying genes and nucleotide polymorphisms causally associated with trait variation. Rigorous standards of proof for determination of the molecular basis of a QTL have been established⁴³, and several recent studies have used methods such as fine mapping and transgenic complementation to identify causal genes^{25,44,45}. However, until homologous recombination techniques for allelic replacement become available for *A. thaliana*⁴⁶, the variation introduced by position effects may limit the power to detect minor allelic effects. This is well illustrated by a study of the fitness effects of inserting a transgene for *Rpm1*-mediated pathogen resistance⁴⁷, where comparisons among five independent insertion lines found that fitness differed by 37% among lines that varied only in the genomic location of the transgene.

What sorts of molecular polymorphisms control quantitative trait variation? Examples from *A. thaliana* span the full range of gene function, including photoreceptor protein polymorphisms⁴⁴, transcription factors⁴⁸, *cis*-regulatory polymorphisms³⁰, insertion/deletion polymorphisms⁴², and copy-number variation in tandem

gene families¹². However, current examples are insufficient to draw generalizations about the importance of regulatory versus coding polymorphisms, especially because the QTL that have been cloned so far have larger than average effects, and may not be representative of small-effect QTL.

Natural selection for complex traits

How does natural selection act on trait variation in the real world? The native range of *A. thaliana* spans a broad range of climates and habitats. *A. thaliana* has evolutionarily expanded its range to warmer climates from the cool temperate 'climate space' inhabited by other *Arabidopsis* species⁴⁹. This climatic range expansion may result from *A. thaliana*'s evolutionary transition to an annual life history from the perennial strategy shown by its close relatives. The annual strategy may increase tolerance of warmer climates by allowing plants to survive summer conditions as seeds rather than as juvenile or adult plants. However, the annual strategy necessitates germinating and flowering during favourable growing conditions, and the seasonal timing of these events varies across the climatic range of the species. Natural populations of *A. thaliana* are thus likely to experience very different selective pressures across the species range. If geographic variation in climate has resulted in local adaptation, we should expect genetic associations between the phenotype of an accession and the climate in its site of origin.

How much natural variation exists for ecologically important traits in *A. thaliana*? Experiments with geographically diverse ecotypes under common garden conditions reveal remarkable genetic variation in many ecologically important traits^{20,23,48,50,51}. Much less is known about the phenology and demography of natural *A. thaliana* populations *in situ*, but it is clear that the species shows variation in life history timing across its range. In particular, there is a life history polymorphism between winter annual ecotypes—which overwinter as small vegetative rosettes and flower in spring—and rapid cycling ecotypes—which germinate and flower in late summer and again in spring to produce two generations per year (ref. 52). This polymorphism has been attributed to loss of a vernalization requirement conferred by mutation at the major flowering time loci, *FRIGIDA* (*FRI*)^{42,53} and *FLOWERING LOCUS C* (*FLC*)⁵⁴, with additional variation due to other loci^{23,48,51}.

Does the observed genetic variation among natural populations represent adaptive differentiation in response to heterogeneous natural selection? This hypothesis can be tested with several different lines of evidence (Table 1). First, is trait variation associated with geographical variation in climate or other ecological factors? Second, is the degree of quantitative variation among populations greater than expected on the basis of differentiation at neutral markers? Third, do local genotypes have higher fitness than foreign genotypes in reciprocal transplant experiments, and if so, is this 'home-court advantage' attributable to selection favouring the local phenotype of the candidate trait? The first two questions have recently been addressed by quantitative genetic studies in common garden environments, and provide some support for the adaptive differentiation hypothesis. However, the third question will require evidence from reciprocal transplant experiments and field studies of selection.

Several studies have detected latitudinal clines in *A. thaliana* growth and morphology, seedling traits, flowering time, vernalization sensitivity, and circadian period^{23,55–58}. Such gradients in complex traits may provide evidence of local adaptation, especially if the direction of the cline is consistent with functional arguments⁵⁸. However, clinal variation may also result from non-adaptive phenomena such as isolation by distance or admixture between two divergent founder populations. Therefore, geographic tests of trait adaptation are most compelling when combined with other forms of evidence.

If population differentiation in a trait is adaptive, the degree of quantitative genetic differentiation in the trait among populations should be greater than the genetic differentiation among populations in neutral molecular markers¹⁸. (Typically, the proportion of genetic variation found among populations for phenotypic traits and neutral molecular markers is quantified by Q_{st} and F_{st} , respectively.) Using this approach, Le Corre¹⁸ measured trait variation within and among natural populations in France. The Q_{st} for bolting time was significantly higher than the overall F_{st} estimated from neutral molecular markers, suggesting adaptive divergence among populations in reproductive timing. Moreover, the F_{st} for functional polymorphism at the principal flowering time gene *FRI* was also significantly higher than F_{st} at marker loci, suggesting adaptive differentiation at that locus. Because *FRI* confers a vernalization requirement, and is known to be associated with flowering time in non-vernalized plants, these

Table 1 | Evidence for adaptation

Method	Known gene?	Advantages	Disadvantages	Reference	Time frame
Transgenic comparison of candidate allele effects	Yes	The gold standard for functional verification and fitness consequences	Transgene position effects Regulatory complications with field trials	47	Current
QTL, RILs, NILs	No	Does not require known gene Modest regulatory requirements	Effects may be confounded with flanking genes	40	Current
Sequence signature	Yes	Capable of detecting very weak selection Does not require phenotypic information	Must control for demographic history Phenotype experiencing selection is unknown	5	Historical
F_{st}/Q_{st}	No	F_{st} can be calculated from relatively few neutral loci Does not require genome-wide molecular markers	Difficult to detect local adaptation when F_{st} is large	18	Historical
Hitchhiking mapping	No	May localize chromosomal region experiencing local adaptation	Phenotype experiencing selection is unknown	75	Historical
Genotype–environment association	No	Can be applied when causal genes are unknown	May be confounded with gene flow or admixture	58	Historical
Multivariate selection analyses	No	Can be applied when causal genes are unknown	May be confounded with unmeasured traits	61	Current
Reciprocal transplants	No	With proper replication, may detect local adaptation Can be applied when causal genes are unknown	Requires access to original populations and environments Further experiments needed to establish mechanism	59	Historical and current

Abbreviations: QTL, quantitative trait loci; RILs, recombinant inbred lines; NILs, near-isogenic lines; F_{st} , the proportion of molecular marker polymorphism found among populations; Q_{st} , the proportion of quantitative genetic variation found among populations.

results suggest that the adaptive divergence of flowering time occurred through selection for loss of *FRI* function in certain populations.

Common garden studies may provide evidence for adaptive population differentiation in complex traits, and suggest hypotheses about selective mechanisms for such differentiation. However, direct support for the local adaptation hypothesis requires reciprocal transplants between natural populations⁵⁹, and selective mechanisms are best tested by measuring natural selection on traits of interest in such experiments. Although this approach has proved to be very powerful in other plant species, it has rarely been attempted for *A. thaliana*, and evidence for local adaptation has been equivocal⁶⁰. A reciprocal-transplant approach could be very valuable for testing hypotheses about local adaptation to climate across the native range of *A. thaliana*.

Although direct experimental tests of local adaptation are still lacking for *A. thaliana*, several field experiments provide strong evidence for real-time natural selection on complex traits and specific loci. Multivariate selection analysis⁶¹ quantifies how natural selection acts on variation in complex traits. Such studies are most valuable when combined with ecological manipulations to test hypotheses about the causes of selection⁶². For example, Donohue *et al.*⁶³ demonstrated geographical differences in selection on germination timing for seeds of the same genotypes experimentally dispersed in Rhode Island and Kentucky, as well as seasonal differences in selection between seeds dispersed in spring and autumn.

How does natural selection on complex traits act on specific loci? Recent field experiments with RILs have identified QTL influencing fitness in different environments^{40,64,65}. The same set of lines showed largely different QTL for fitness in Georgia⁴⁰, North Carolina, and Rhode Island⁶⁵, suggesting that natural selection acts on different genomic regions in different geographic locations and seasonal environments. These studies also revealed pervasive epistatic selection. For example, Malmberg and colleagues⁴⁰ found epistatic QTL to be more common and more important than non-epistatic QTL, and found a high frequency of sign epistasis for fecundity. Such epistatic interactions can maintain genetic variation and constrain evolutionary responses to natural selection³⁶, and may provide a possible mechanism for the maintenance of polymorphism in *A. thaliana*.

These studies could not test hypotheses about local adaptation, as the parents of the RILs were not native to the field sites. Rather, they provide insights into how natural selection will act on the segregating variation in different novel environments, a potentially relevant scenario for this colonizing species. Such experiments also can test whether fitness QTL co-localize with QTL for complex traits experiencing natural selection, thus providing insights into selective mechanisms⁶⁴. The next step is to take RILs from locally adapted parental ecotypes and conduct selection experiments with the RILs in the parental sites of origin. Such experiments can elucidate the selective mechanisms and specific loci contributing to local adaptation⁶⁶.

Near-isogenic lines (NILs) containing alternate alleles of polymorphic candidate genes provide another promising tool for testing hypotheses about how selection acts on polymorphisms in ecological time. Such studies will be particularly useful for polymorphisms that show the molecular signature of balancing selection in genes of known ecological function. For example, *R*-genes in *A. thaliana* confer resistance to pathogens, and are often polymorphic for resistant or susceptible alleles, with evidence for balancing selection at the molecular level^{13,67,68}. Experiments with NILs⁶⁹ and pairs of transgenics that differed in the presence or absence of a resistant allele⁴⁷ were able to demonstrate fitness costs of resistance alleles, which provide an ecological mechanism for the maintenance of these polymorphisms. This approach may be valuable for future studies of the selective mechanisms acting to maintain variation at other candidate loci.

Future prospects

Future advances in our understanding of *A. thaliana* natural variation depend, in part, on improvements in resources and technology.

Among the highest priorities are expanded collections from well-characterized environments, as well as possible undisturbed populations from Eurasia. On the technical side, the greatest need is for straightforward protocols to reduce transgene position effects, such as gene targeting⁴⁶ or alternative approaches^{47,70}.

Arabidopsis research is poised to address fundamental questions in evolution and ecology, including: What is the role of epistasis, and how does it influence the evolution of developmental pathways? How do ecological processes influence balancing selection within and among populations? What are the frequencies and effects of QTL alleles, and how is this influenced by evolutionary forces? What are the ecological and molecular mechanisms of local adaptation within the native range of *A. thaliana*? How do real-time ecological processes influence adaptive evolution in introduced and invasive populations? As the field matures, research will focus increasingly on related species, which will greatly improve our understanding of the genetic basis and evolutionary mechanisms of speciation.

Equally as important, *Arabidopsis* biology encourages (and requires) interdisciplinary research and training. A new synthesis of functional genomics with evolution and ecology will benefit each component discipline, and will bring fundamental advances in our understanding of biological mechanisms and processes. Such training provides an essential component of science education for the future.

- Alonso-Blanco, C., Mendez-Vigo, B. & Koornneef, M. From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *Int. J. Dev. Biol.* **49**, 717–732 (2005).
- Hoffmann, M. H. Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). *J. Biogeogr.* **29**, 125–134 (2002).
- Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, e196 (2005).
- Schmid, K. J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B. & Mitchell-Olds, T. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* **169**, 1601–1615 (2005).
- Wright, S. I. & Gaut, B. S. Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.* **22**, 506–519 (2005).
- Ometto, L., Glinka, S., De Lorenzo, D. & Stephan, W. Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Mol. Biol. Evol.* **22**, 2119–2130 (2005).
- Akey, J. M. *et al.* Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* **2**, e286 (2004).
- Abbott, R. J. & Gomes, M. F. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62**, 411–418 (1989).
- Bustamante, C. D. *et al.* The cost of inbreeding in *Arabidopsis*. *Nature* **416**, 531–534 (2002).
- Shimizu, K. K. *et al.* Darwinian selection on a selfing locus. *Science* **306**, 2081–2084 (2004).
- Cork, J. M. & Purugganan, M. D. High-diversity genes in the *Arabidopsis* genome. *Genetics* **170**, 1897–1911 (2005).
- Kroymann, J., Donnerhacke, S., Schnabelrauch, D. & Mitchell-Olds, T. Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proc. Natl Acad. Sci. USA* **100**, 14587–14592 (2003).
- Tian, D., Araki, H., Stahl, E., Bergelson, J. & Kreitman, M. Signature of balancing selection in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 11525–11530 (2002).
- Turelli, M. & Barton, N. H. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G × E interactions. *Genetics* **166**, 1053–1079 (2004).
- Schmid, K. *et al.* Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theor. Appl. Genet.* **112**, 1104–1114 (2006).
- Jorgensen, S. & Mauricio, R. Neutral genetic variation among wild North American populations of the weedy plant *Arabidopsis thaliana* is not geographically structured. *Mol. Ecol.* **13**, 3403–3413 (2004).
- Bakker, E. G. *et al.* Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. *Mol. Ecol.* **15**, 1405–1418 (2006).
- Le Corre, V. Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Mol. Ecol.* **14**, 4181–4192 (2005).
- Stenoien, H. K., Fenster, C. B., Tonteri, A. & Savolainen, O. Genetic variability in natural populations of *Arabidopsis thaliana* in northern Europe. *Mol. Ecol.* **14**, 137–148 (2005).
- Koornneef, M., Alonso-Blanco, C. & Vreugdenhil, D. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* **55**, 141–172 (2004).

21. Bouche, N. & Bouchez, D. *Arabidopsis* gene knockout: phenotypes wanted. *Curr. Opin. Plant Biol.* **4**, 111–117 (2001).
22. Schmid, M. *et al.* A gene expression map of *Arabidopsis thaliana* development. *Nature Genet.* **37**, 501–506 (2005).
23. Lempe, J. *et al.* Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genetics* **1**, e6 (2005).
24. Mackay, T. F. C. The genetic architecture of quantitative traits: lessons from *Drosophila*. *Curr. Opin. Genet. Dev.* **14**, 253–257 (2004).
25. Kroymann, J. & Mitchell-Olds, T. Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435**, 95–98 (2005).
26. Fishman, L., Kelly, A. & Willis, J. Minor quantitative trait loci underlie floral traits associated with mating system divergence in *Mimulus*. *Evolution* **56**, 2138–2155 (2002).
27. Gachon, C. M. M., Langlois-Meurinne, M., Henry, Y. & Saindrenan, P. Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: functional and evolutionary implications. *Plant Mol. Biol.* **58**, 229–245 (2005).
28. Kliebenstein, D. J. *et al.* Genomic survey of gene expression diversity in *Arabidopsis thaliana*. *Genetics* **172**, 1179–1189 (2006).
29. Vuytsteke, M., van Eeuwijk, F., Van Hummelen, P., Kuiper, M. & Zabeau, M. Genetic analysis of variation in gene expression in *Arabidopsis thaliana*. *Genetics* **171**, 1267–1275 (2005).
30. DeCook, R., Lall, S., Nettleton, D. & Howell, S. H. Genetic regulation of gene expression during shoot development in *Arabidopsis*. *Genetics* **172**, 1155–1164 (2006).
31. de Meaux, J., Goebel, U., Pop, A. & Mitchell-Olds, T. Allele-specific assay reveals functional variation in the *Chalcone Synthase* promoter of *Arabidopsis thaliana* that is compatible with neutral evolution. *Plant Cell* **17**, 676–690 (2005).
32. Gibson, G. & Weir, B. The quantitative genetics of transcription. *Trends Genet.* **21**, 616–623 (2005).
33. Ungerer, M. C., Halldorsdottir, S. S., Purugganan, M. D. & Mackay, T. F. C. Genotype–environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. *Genetics* **165**, 353–365 (2003).
34. Juenger, T. E., Sen, S., Stowe, K. A. & Simms, E. L. Epistasis and genotype–environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica* **123**, 87–105 (2005).
35. Dilda, C. L. & Mackay, T. F. C. The genetic architecture of *Drosophila* sensory bristle number. *Genetics* **162**, 1655–1674 (2002).
36. Weinreich, D., Watson, R. & Chao, L. Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
37. Syed, N. H. & Chen, Z. J. Molecular marker genotypes, heterozygosity and genetic interactions explain heterosis in *Arabidopsis thaliana*. *Heredity* **94**, 295–304 (2004).
38. Hausmann, N. J. *et al.* Quantitative trait loci affecting delta C-13 and response to differential water availability in *Arabidopsis thaliana*. *Evolution* **59**, 81–96 (2005).
39. Kearsley, M., Pooni, H. & Syed, N. Genetics of quantitative traits in *Arabidopsis thaliana*. *Heredity* **91**, 456–464 (2003).
40. Malmberg, R. L., Held, S., Waits, A. & Mauricio, R. Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. *Genetics* **171**, 2013–2027 (2005).
41. Michaels, S. D. & Amasino, R. M. Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell* **13**, 935–942 (2001).
42. Johanson, U. *et al.* Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**, 344–347 (2000).
43. Weigel, D. & Nordborg, M. Natural variation in *Arabidopsis*. How do we find the causal genes? *Plant Physiol.* **138**, 567–568 (2005).
44. El-Assal, S. E.-D., Alonso-Blanco, C., Peeters, A. J. M., Raz, V. & Koornneef, M. A. QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genet.* **29**, 435–440 (2001).
45. Werner, J. D. *et al.* Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. *Proc. Natl Acad. Sci. USA* **102**, 2460–2465 (2005).
46. Puchta, H. & Hohn, B. Green light for gene targeting in plants. *Proc. Natl Acad. Sci. USA* **102**, 11961–11962 (2005).
47. Tian, D., Traw, M., Chen, J., Kreitman, M. & Bergelson, J. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74–77 (2003).
48. Werner, J. D. *et al.* *FRIGIDA*-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. *Genetics* **170**, 1197–1207 (2005).
49. Hoffmann, M. H. Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). *Evolution* **59**, 1425–1436 (2005).
50. Pigliucci, M. In *The Arabidopsis Book* (eds Somerville, C. R. & Meyerowitz, E. M.) (American Society of Plant Biologists, Rockville, Maryland, 2002) doi:10.1199/tab.0009 (www.aspb.org/publications/arabidopsis/) (2002).
51. Shindo, C. *et al.* Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. *Plant Physiol.* **138**, 1163–1173 (2005).
52. Thompson, L. The spatiotemporal effects of nitrogen and litter on the population dynamics of *Arabidopsis thaliana*. *J. Ecol.* **82**, 63–68 (1994).
53. Simpson, G. G. & Dean, C. *Arabidopsis*, the Rosetta stone of flowering time? *Science* **296**, 285–289 (2002).
54. Michaels, S. D., He, Y., Scortecci, K. C. & Amasino, R. M. Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **100**, 10102–10107 (2003).
55. Maloof, J. N. *et al.* Natural variation in light sensitivity of *Arabidopsis*. *Nature Genet.* **29**, 441–446 (2001).
56. Michael, T. P. *et al.* Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science* **302**, 1049–1053 (2003).
57. Stenoien, H. K., Fenster, C. B., Kuittinen, H. & Savolainen, O. Quantifying latitudinal clines to light responses in natural populations of *Arabidopsis thaliana* (Brassicaceae). *Am. J. Bot.* **89**, 1604–1608 (2002).
58. Stinchcombe, J. R. *et al.* A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proc. Natl Acad. Sci. USA* **101**, 4712–4717 (2004).
59. Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241 (2004).
60. Callahan, H. S. & Pigliucci, M. Shade-induced plasticity and its ecological significance in wild populations of *Arabidopsis thaliana*. *Ecology* **83**, 1965–1980 (2002).
61. Rauscher, M. D. The measurement of selection on quantitative traits—biases due to environmental covariances between traits and fitness. *Evolution* **46**, 616–626 (1992).
62. Wade, M. J. & Kalisz, S. The causes of natural selection. *Evolution* **44**, 1947–1955 (1990).
63. Donohue, K. *et al.* The evolutionary ecology of seed germination of *Arabidopsis thaliana*: variable natural selection on germination timing. *Evolution* **59**, 758–770 (2005).
64. Callahan, H. S., Dhanooolal, N. & Ungerer, M. C. Plasticity genes and plasticity costs: a new approach using an *Arabidopsis* recombinant inbred population. *New Phytol.* **166**, 129–139 (2005).
65. Weinig, C. *et al.* Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics* **165**, 321–329 (2003).
66. Verhoeven, K. J. F., Vanhala, T. K., Biere, A., Nevo, E. & Van Damme, J. M. M. The genetic basis of adaptive population differentiation: a quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. *Evolution* **58**, 270–283 (2004).
67. Mauricio, R. *et al.* Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. *Genetics* **163**, 735–746 (2003).
68. Shen, J., Araki, H., Chen, L., Chen, J.-Q. & Tian, D. Unique evolutionary mechanism in R-genes under the presence/absence polymorphism in *Arabidopsis thaliana*. *Genetics* **172**, 1243–1250 (2006).
69. Korves, T. & Bergelson, J. A novel cost of R gene resistance in the presence of disease. *Am. Nat.* **163**, 489–504 (2004).
70. Siegal, M. L. & Hartl, D. L. Transgene coplacement and high efficiency site-specific recombination with the Cre/loxP system in *Drosophila*. *Genetics* **144**, 715–726 (1996).
71. Plagnol, V., Padhukasahasram, B., Wall, J. D., Marjoram, P. & Nordborg, M. Relative influences of crossing-over and gene conversion on the pattern of linkage disequilibrium in *Arabidopsis thaliana*. *Genetics* **172**, 2441–2448 (2005).
72. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genet.* **38**, 203–208 (2006).
73. Remington, D. L. *et al.* Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl Acad. Sci. USA* **98**, 11479–11484 (2001).
74. Aranzana, M. *et al.* Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics* **1**, e60 (2005).
75. Schlottner, C. Hitchhiking mapping—functional genomics from the population genetics perspective. *Trends Genet.* **19**, 32–38 (2003).

Acknowledgements We thank M. Noor, M. Nordborg and our collaborators and laboratory members for discussion and comments. T.M.-O. and J.S. were supported by Duke University and the National Science Foundation, respectively.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to T.M.-O. (tmo1@duke.edu).

The magnetic nature of disk accretion onto black holes

Jon M. Miller¹, John Raymond², Andy Fabian³, Danny Steeghs², Jeroen Homan⁴, Chris Reynolds⁵, Michiel van der Klis⁶ & Rudy Wijnands⁶

Although disk accretion onto compact objects—white dwarfs, neutron stars and black holes—is central to much of high-energy astrophysics, the mechanisms that enable this process have remained observationally difficult to determine. Accretion disks must transfer angular momentum in order for matter to travel radially inward onto the compact object¹. Internal viscosity from magnetic processes^{1–4} and disk winds⁵ can both in principle transfer angular momentum, but hitherto we lacked evidence that either occurs. Here we report that an X-ray-absorbing wind discovered in an observation of the stellar-mass black hole binary GRO J1655–40 (ref. 6) must be powered by a magnetic process that can also drive accretion through the disk. Detailed spectral analysis and modelling of the wind shows that it can only be powered by pressure generated by magnetic viscosity internal to the disk or magnetocentrifugal forces. This result demonstrates that disk accretion onto black holes is a fundamentally magnetic process.

To study the nature of disk accretion onto black holes, we observed the transient source GRO J1655–40 with the Chandra X-ray Observatory for 63.5 ks starting on 1 April 2005 at 12:41:44 (terrestrial time, TT), during an X-ray bright phase of its 2005 outburst. GRO J1655–40 is a binary system at a distance of 3.2 kpc that harbours a black hole with a mass of $7.0M_{\odot}$ ($M_{\odot}$ is the solar mass), which accretes from an F3 IV–F6 IV star with a mass of $2.3M_{\odot}$ in a 2.6-day orbit⁶. The inner disk is viewed at an inclination of $67\text{--}85^{\circ}$ (nearly edge-on)^{6,7}. Using Chandra, we obtained a robust high-resolution X-ray spectrum of GRO J1655–40 in the soft X-ray band (Fig. 1).

It is common to decompose the broad-band spectra of stellar-mass black holes into disk blackbody and power-law components. Assuming the standard equivalent neutral hydrogen absorption⁸ along the line of sight to GRO J1655–40 (equivalent neutral hydrogen column density $N_{\text{H}} = 7.4 \times 10^{21} \text{ atoms cm}^{-2}$), this spectral model gives a disk temperature of $kT = 1.34(1) \text{ keV}$ (where k is Boltzmann's constant, and numbers in parentheses indicate the error in the last significant digit) and a photon power-law index of $\Gamma = 3.54(1)$ in the 1.2–19 Å range, and a total flux of $4.70(5) \times 10^{-8} \text{ erg cm}^{-2} \text{ s}^{-1}$ (unabsorbed). This flux implies a luminosity of $L = 3.3 \times 10^{37} \text{ erg s}^{-1}$ for a distance from Earth of $d = 3.2 \text{ kpc}$, or 4% of the Eddington limit for a $7M_{\odot}$ black hole. The disk contributes 65% of the unabsorbed flux.

The HETGS (High Energy Transmission Grating Spectrometer) spectra of GRO J1655–40 contain 90 absorption lines significant at the 5σ level of confidence or higher. We can confidently identify 76 of these lines with resonance lines expected from over 32 charge states. The properties of these lines were measured with simple gaussian line functions and local continuum models. Line centroid

wavelengths and oscillator strengths were taken from a set of standard references^{9–11}. Two findings demand that the absorption arises in a disk-driven wind. First, the lines show blueshifts in the $300\text{--}1,600 \text{ km s}^{-1}$ range, indicating a flow into our line of sight (Figs 1 and 2, and Supplementary Information). Second, the spectra contain no strong emission lines; this fact signals that the absorbing gas is mostly equatorial along the plane of the disk.

To understand better the nature of the wind, we constructed a number of photoionized plasma models based on the methodology and atomic physics packages described in earlier work^{12,13}. (The models used in this work differ only in that new dielectronic recombination rates for Fe and Ni were used to more accurately describe L-shell ions¹⁴.) These models describe a gas in photoionization equilibrium, with heating by photoionization and Compton

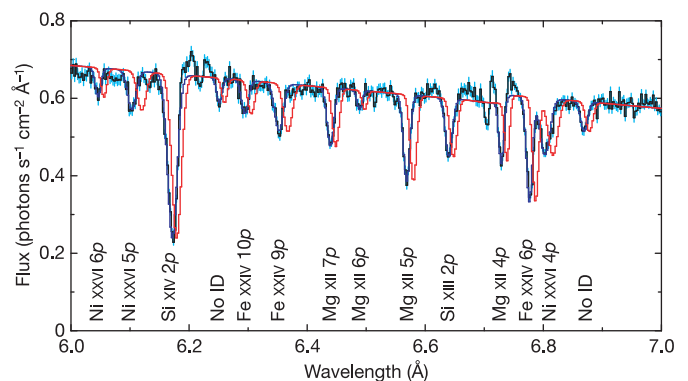


Figure 1 | A small part of the disk wind spectrum observed in GRO J1655–40 with Chandra. The best-fit phenomenological model is shown in blue. The model in red plots the natural line wavelengths, to illustrate that the observed absorption is blueshifted. The errors shown are 1σ statistical errors on the photon flux. To obtain this high-resolution spectrum, the Chandra HETGS was used to disperse the X-ray flux onto the Advanced CCD Imaging Spectrometer (ACIS), which was operated in ‘continuous-clocking’ mode. The data reduction and preparation were performed using the latest version of the standard Chandra packages (CIAO), and a procedure typical for this instrumental configuration¹³. The spectra presented in this work were produced by adding the first-order spectra from the HETGS medium-energy grating (MEG) at a resolution of 0.005 Å per bin, and the first-order spectra from the high-energy grating (HEG) at a resolution of 0.0025 Å per bin. The HEG spectrum was used to characterize the 1–11 Å band, and the MEG spectrum was used to characterize the 11–19 Å band. The continuum emission was characterized using the HEG spectrum. All spectral fits were made using the ISIS²⁹ spectral fitting package.

¹Department of Astronomy, University of Michigan, 500 Church Street, Ann Arbor, Michigan 48109, USA. ²Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ³Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK. ⁴Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. ⁵Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. ⁶Astronomical Institute Anton Pannekoek, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands.

scattering and cooling by line and continuum emission and Compton scattering. The predicted ionization state of the gas is combined with a curve of growth¹⁵ for a chosen velocity width to determine line equivalent widths using Voigt profiles. The code accounts for line blends self-consistently. A standard set of elemental abundances are used in the code¹⁶. The illuminating spectrum is taken to be the composite thermal and non-thermal continuum spectrum described briefly above.

The observed spectrum is consistent with absorption in a constant-density slab with a thickness of approximately 2.5×10^8 cm and a number density of $n = 5.6 \times 10^{15}$ atoms cm^{-3} at a mean distance of 4.8×10^8 cm from the black hole. This translates to approximately $200R_{\text{Schw}}$ (where the Schwarzschild radius $R_{\text{Schw}} = 2GM_{\text{BH}}/c^2$, with G the gravitational constant, M_{BH} the black hole mass, and c the velocity of light); the component of the wind velocity in our line of sight is much slower than the local virial velocity. The temperature of the gas is $(0.2\text{--}1.0) \times 10^6$ K. An intrinsic full-width at half-maximum (FWHM) of 300 km s^{-1} matches the observed lines which are not saturated (Fig. 2 and Supplementary Information). An important feature of this wind is that it is highly ionized: $\log \xi = \log(L_{\text{X}}/nr^2) = 4.2\text{--}4.7$ (where ξ is the ionization parameter, L_{X} is X-ray luminosity, n is density, and r is the radius from the ionizing source).

The data permit strong independent constraints on the geometry and extent of the wind absorption. An inner radial extent of $10^{7.5}$ cm is set by the density at which Fe XXIII lines are produced. At smaller radii, the ionization parameter requires a density so large that the metastable $2s2p^3$ P level of Fe XXIII would be populated, producing lines that are not observed. An outer radial extent of $10^{9.5}$ cm is set by dilution and line ratios. At larger radii, the thickness of the absorbing gas becomes large compared to the distance, so $1/r^2$ dilution of the radiation field makes it impossible to get enough total column at high enough ionization parameter. A lower limit of 6° on the height of the gas above the disk midplane comes from the fact that the line of sight must pass over the outer edge of the disk¹⁷. An upper limit on the vertical extent comes from the lack of emission features. A Monte Carlo simulation for a cascade resulting from absorption in the $2s\text{--}4p$ line of Fe XXIV predicts a $2p\text{--}3s$ emission line with equivalent width

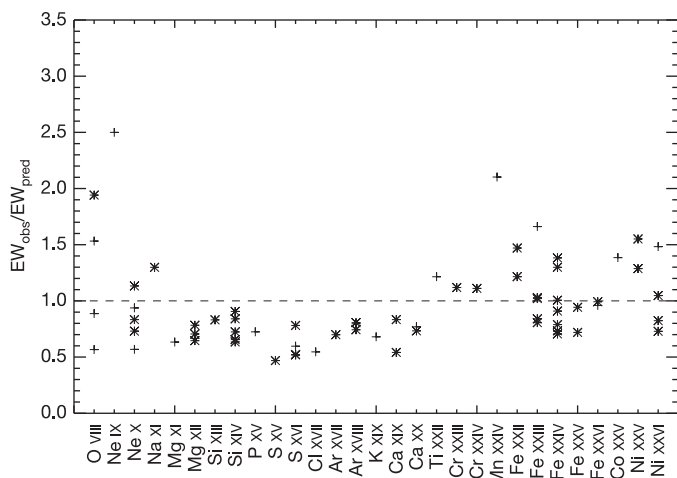


Figure 2 | Comparison to the data of the best model for the disk wind in GRO J1655–40. The plot shows the ratio of the absorption line equivalent widths measured in GRO J1655–40, EW_{obs} , to the equivalent widths predicted by the photoionization model that best describes the disk wind, EW_{pred} . The model assumes internal velocity widths of 300 km s^{-1} . Revised solar abundances¹⁶ are adequate to describe the elements between Na and K (inclusive). Abundances of twice the revised solar value are required to describe the lines observed from other elements. For visual clarity, stronger lines are marked with stars, and weaker lines with crosses. The full spectrum of GRO J1655–40 and the measured properties of the absorption lines are detailed in the Supplementary Information.

$0.65 \times 28\Omega/4\pi = 18\Omega/4\pi \text{ mÅ}$ (where Ω is the solid angle covering factor). An upper limit of a 2 mÅ to such a feature implies an upper limit of about 12° to the vertical extent of the absorbing gas.

The properties of thermally driven winds have been studied, especially within the context of outer accretion disks being illuminated by a central X-ray source¹⁸. In such cases, it is possible to define a critical radius $R_C = (1.0 \times 10^{10}) \times (M_{\text{BH}}/M_\odot)/(T_{\text{C8}}) \text{ cm}$, and a wind may occur for any $R/R_C > 0.1$ (where T_{C8} is the gas temperature in units of 10^8 K). On the basis of temperatures derived from our photoionization models, the smallest possible value of R_C for the wind observed in GRO J1655–40 is 7×10^{12} cm. Thus, the minimum radius at which a disk wind can be thermally driven is approximately two orders of magnitude greater than is plausible in GRO J1655–40. Similar results are obtained when our results are compared to new models for the winds in active galactic nuclei (AGN) such as NGC 3783 (ref. 19).

The wind observed is too highly ionized to be driven by radiation pressure. The overwhelming majority of absorption lines observed are from He-like and H-like species, which means that there is little ultraviolet opacity in the wind by which momentum may be transferred. At ionization parameters of $\xi > 10^3$, models for line-driven winds in AGN indicate that ultraviolet emission lines provide no additional driving force²⁰. Moreover, our photoionization code measures radiation pressure as the momentum of the photons absorbed, which is comparable to the electron scattering radiation pressure, and even together these effects fall far short of producing the observed momentum flux in the wind.

Given that thermal and radiative driving fail by orders of magnitude, magnetic driving of the wind in GRO J1655–40 is the only plausible mechanism remaining. Our model implies a mass loss rate in the wind of $\dot{m}_w = 3.5 \times 10^{17} \text{ g s}^{-1}$ or approximately $0.5 \text{ g cm}^{-2} \text{ s}^{-1}$. For a typical blueshift of 500 km s^{-1} , this translates into a kinetic energy flux of $6.3 \times 10^{14} \text{ erg cm}^{-2} \text{ s}^{-1}$. An angle of 9° above the disk midplane at a radius of 4.8×10^8 cm from the black hole corresponds to a height of $Z = 0.15r$; an energy flux of $2.0 \times 10^{16} \text{ erg cm}^{-2} \text{ s}^{-1}$ is required to lift the gas to that height. The luminosity of the central engine as derived from fits to the continuum implies a mass accretion rate of $\dot{m}_a = 3.7 \times 10^{17} \text{ g s}^{-1}$ for an accretion efficiency of 10%. The viscous energy flux dissipated is given by $3GM_{\text{BH}}\dot{m}_a/4\pi r^3 = 2.3 \times 10^{18} \text{ erg cm}^{-2} \text{ s}^{-1}$. Recent simulations show that the magnetorotational instability^{2,3,4} can not only drive turbulence, viscosity and accretion through a disk, but can transmit 25% of the magnetic energy flux vertically out of the disk²¹ and drive a wind. In the case of GRO J1655–40, 25% of the viscous energy flux is comparable to the flux required to drive the wind to infinity. The wind outflow velocity and mass outflow rate are remarkably similar to predictions resulting from simulations of magnetically driven winds from magneto-rotational instability (MRI) disks²².

It is also possible that the wind is driven by magnetocentrifugal forces⁵. The absorption spectrum contains no information about velocities tangential to our line of sight. Theoretical models show that largely equatorial winds can arise via magnetocentrifugal driving when the magnetic field vector makes a small angle to the disk plane²³; however, recent simulations suggest that it is difficult to maintain the poloidal magnetic field required in a wind with a large mass outflow²². Given that the wind arises in a disk that is almost certainly keplerian, however, magnetocentrifugal driving cannot be discounted. The winds in some young stars (FU Orionis class) are driven by magnetocentrifugal forces²⁴. Tapping into the rotational velocity of the disk would help to expel the wind to infinity. Although internal magnetic viscosity and magnetic winds are sometimes posited as complete and separate processes, the most physically realistic scenario may be one in which both processes act to drive disk accretion and outflows.

Winds are commonly observed in accreting compact objects; however, GRO J1655–40 is the first case wherein it is clear that the

wind must be launched from the disk (not the companion star) and must be driven primarily by magnetic processes (not thermal and radiative pressure). An X-ray wind with some similarities to those in AGN was detected in the accreting neutron star binary Circinus X-1 (ref. 25); however, in Circinus X-1 contributions from a massive companion star cannot be ruled out, and the wind could be driven by thermal and radiation pressure²⁵. A wind more certainly tied to the disk was detected in the black hole binary GX 339–4, but too few lines were detected to constrain the driving mechanism¹³. Though radiative driving may be important in most white dwarf systems, there is at least one system where it may be inadequate²⁶. Similarly, theoretical considerations suggest that magnetocentrifugal forces are important in AGN winds^{27,28}, but present data do not yet require or rule out these effects. Our results therefore represent a rare and crucial insight into the nature of the processes that drive disk accretion in black holes.

Magnetic pressure supplied by the disk provides a natural means of driving the highly ionized winds observed in many accreting stellar-mass black holes and neutron stars, and may play a role in driving the hottest winds observed in AGN. Indeed, winds and jets are ubiquitous features in accretion-powered astrophysics, and the role of magnetic processes revealed in GRO J1655–40 gives a broad observational insight into the physical coupling between inflows and outflows in accreting compact objects.

Received 22 December 2005; accepted 10 May 2006.

- Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).
- Balbus, S. A. & Hawley, J. F. A powerful local shear instability in weakly magnetized disks. *Astrophys. J.* **376**, 214–233 (1991).
- Hawley, J. F., Gammie, C. F. & Balbus, S. A. Local three-dimensional magneto-hydrodynamic solutions of accretion disks. *Astrophys. J.* **440**, 742–763 (1995).
- Balbus, S. A. & Hawley, J. F. Instability, turbulence, and enhanced transport in accretion disks. *Rev. Mod. Phys.* **70**, 1–53 (1998).
- Blandford, R. D. & Payne, D. G. Hydromagnetic flows from accretion disks and the production of radio jets. *Mon. Not. R. Astron. Soc.* **199**, 883–903 (1982).
- Orosz, J. & Bailyn, C. D. Optical observations of GRO J1655–40 in quiescence. I. A precise mass for the black hole primary. *Astrophys. J.* **477**, 876–896 (1997).
- Hjellming, R. M. & Rupen, M. P. Episodic ejection of relativistic jets by the X-ray transient GRO J1655–40. *Nature* **375**, 464–468 (1995).
- Dickey, J. M. & Lockman, F. J. H I in the Galaxy. *Annu. Rev. Astron. Astrophys.* **28**, 215–261 (1990).
- Verner, D. A., Verner, E. M. & Ferland, G. J. Atomic data for permitted resonance lines of atoms and ions from H to Si, and S, Ar, Ca, and Fe. *Atom. Data Nucl. Data Tables* **64**, 11–180 (1996).
- The NIST Atomic Spectra Database, Standard Reference Database 78. (<http://physics.nist.gov/cgi-bin/AtData/main.asd>) (2005).
- Nahar, S. & Pradhan, A. K. Atomic data from the Iron Project. XXXV. Relativistic fine structure oscillator strengths for Fe XXIV and Fe XXV. *Astron. Astrophys. Suppl.* **135**, 347–357 (1999).
- Raymond, J. A model of an X-ray-illuminated accretion disk and corona. *Astrophys. J.* **412**, 267–277 (1993).
- Miller, J. M. et al. Chandra/HETGS spectroscopy of the Galactic black hole GX 339–4: A relativistic iron emission line and evidence for a Seyfert-like warm absorber. *Astrophys. J.* **601**, 450–465 (2004).
- Colgan, J., Pindzola, M. S. & Badnell, N. R. Dielectronic recombination data for dynamic finite-density plasmas. V: the lithium isoelectronic sequence. *Astron. Astrophys.* **417**, 1183–1188 (2004).
- Spitzer, L. *Physical Processes in the Interstellar Medium* (Wiley, New York, 1978).
- Grevesse, N. & Sauval, A. J. in *Solar Composition and its Evolution—from Core to Corona* (eds Frölich, C., Huber, M. C. E., Solanski, S. K. & von Steiger, R.) 161–174 (Kluwer, Dordrecht, 1998).
- Vrtilek, S. et al. Observations of Cygnus X-2 with IUE—Ultraviolet results from a multi-wavelength campaign. *Astron. Astrophys.* **234**, 162–173 (1990).
- Begelman, M. C., McKee, C. F. & Shields, G. A. Compton heated winds and coronae above accretion disks. II Dynamics. *Astrophys. J.* **271**, 70–89 (1983).
- Chelouche, D. & Netzer, H. Dynamical and spectral modeling of the ionized gas and nuclear environment in NGC 3783. *Astrophys. J.* **625**, 95–107 (2005).
- Proga, D., Stone, J. M. & Kallman, T. R. Dynamics of line-driven winds in active galactic nuclei. *Astrophys. J.* **543**, 686–696 (2000).
- Miller, K. A. & Stone, J. M. The formation and structure of a strongly magnetized corona above a weakly magnetized accretion disk. *Astrophys. J.* **534**, 398–419 (2000).
- Proga, D. Numerical simulations of mass outflows driven from accretion disks by radiation and magnetic forces. *Astrophys. J.* **585**, 406–417 (2003).
- Spruit, H. C. in *Physical Processes in Binary Stars* (eds Wijers, R. A. M. J., Davies, M. B. & Tout, C. A.) 249–286 (NATO ASI Ser., Kluwer, Dordrecht, 1996).
- Calvet, N., Hartmann, L. & Kenyon, S. J. Mass loss from pre-main-sequence accretion disks. I—The accelerating wind of FU Orionis. *Astrophys. J.* **402**, 623–634 (1993).
- Schulz, N. S. & Brandt, W. N. Variability of the X-ray P Cygni line profiles from Circinus X-1 near zero phase. *Astrophys. J.* **572**, 972–983 (2002).
- Mauche, C. W. & Raymond, J. C. Extreme Ultraviolet Explorer observations of OY Carinae in superoutburst. *Astrophys. J.* **541**, 924–936 (2000).
- Königl, A. & Kartje, J. F. Disk-driven hydromagnetic winds as a key ingredient of active galactic nuclei unification schemes. *Astrophys. J.* **434**, 446–467 (1994).
- Everett, J. E. Radiative transfer and acceleration in magnetocentrifugal winds. *Astrophys. J.* **631**, 689–706 (2005).
- Houck, J. C. & Denicola, L. A. ISIS: An interactive spectral interpretation system for high resolution X-ray spectroscopy. *Astron. Soc. Pacif. Conf. Proc.* **216**, 591–594 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge conversations with N. Calvet, L. Hartmann, D. Proga and M. Rupen. We are indebted to A. Prestwich, H. Tananbaum and the Chandra staff for help in making this observation possible. We thank B. Lauritsen for editorial insights. This work was supported by NASA through the Chandra guest observer programme (J.M.M.).

Author Contributions J.M.M. analysed the Chandra data and wrote most of the paper. J.R. developed the photoionization model. J.M.M., J.R., A.F. and C.R. developed the interpretation of the data. D.S., J.H., M.K. and R.W. contributed insights on X-ray binaries and/or made supporting observations with other instruments. All others discussed the work at length, and contributed to the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M.M. (jonmm@umich.edu).

LETTERS

The giant electromechanical response in ferroelectric relaxors as a critical phenomenon

Z. Kutnjak¹, J. Petzelt² & R. Blinc¹

The direct conversion of electrical energy to mechanical work by a material is relevant to a number of applications. This is illustrated by ferroelectric 'relaxors'^{1–4} such as $\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3\text{-PbTiO}_3$ (PMN-PT; refs 5, 6): these materials exhibit a giant electromechanical (piezoelectric) response that is finding use in ultrasonic⁴ and medical applications, as well as in telecommunications. The origins of this effect are, however, still unclear. Here we show that the giant electromechanical response in PMN-PT (and potentially other ferroelectric relaxors) is the manifestation of critical points that define a line in the phase diagram of this system. Specifically, in the electric-field–temperature–composition phase diagram of PMN-PT (the composition being varied by changing the PT concentration), a first-order paraelectric–ferroelectric phase transition terminates in a line of critical points where the piezoelectric coefficient is maximum. Above this line, supercritical evolution is observed. On approaching the critical point, both the energy cost and the electric field necessary to induce ferroelectric polarization rotations decrease significantly, thus explaining the giant electromechanical response of these relaxors.

The perovskite solid solution PMN (refs 5, 6), generally viewed as a canonical relaxor, is characterized by site and charge disorder, which retains its average cubic symmetry down to temperatures as low as 5 K in a state similar to those observed in dipolar glasses^{1–3,7}. But in contrast to dipolar glasses, the ferroelectric state in relaxor systems can be induced by sufficiently high electric fields, thus indicating that these systems must be close to some criticality^{1,2}. However, so far the existence of such a criticality has not been confirmed nor its nature explored. Similarly to the application of an electric field, the addition of PT to PMN also induces ferroelectric behaviour⁵. The PMN-PT system is at low temperatures rhombohedral up to some specific PT content. At still higher PT concentrations, it undergoes a morphotropic phase transition and becomes tetragonal. Giant piezoelectric response is observed near this transition. At higher temperatures, the system becomes cubic for all PT concentrations.

The giant piezoelectric and dielectric responses of PMN-PT-type relaxor materials near the morphotropic phase boundary have been in the past related to reorientation of polar domains by the applied electric field and/or to field-induced polarization rotations^{8–12}. First-principles studies of a BaTiO_3 single domain crystal at zero temperature⁹ have indeed shown that polarization rotations (that is, the motion of the central Ti ion among different equilibrium sites) and the resulting unit cell deformation and elastic strain can account for the above effects.

So far it has been generally accepted that in the two-dimensional phase diagram of the PMN-PT system—temperature (T) versus PT concentration (x), with a zero electric field, E —the giant electromechanical response is largest at the morphotropic phase boundary.

Here we show that in the third (E) dimension there is a new phenomenon, namely a line of critical end points of the liquid–vapour type not previously known in relaxor ferroelectrics. The large polarization fluctuations in the vicinity of this line are responsible for the giant electromechanical response. Our experimental data demonstrate that in the E – T – x phase diagram of PMN-PT the piezoelectric response and static dielectric constant always exhibit a maximum at the critical electric field E_{CL} corresponding to the line of critical end points. The electric field necessary for polarization rotations and the energy barrier involved both significantly decrease at $E = E_{\text{CL}}$. This reveals a new driving mechanism for polarization rotations and the giant electromechanical response of relaxors, and should be helpful in searching for new materials with ultrahigh performance.

We have investigated single crystals of PMN-PT with PT concentrations $x = 0, 0.10, 0.15, 0.25$ and 0.295 . We used a.c. dielectric spectroscopy (20 Hz–1 MHz), quasi-static dielectric and spontaneous polarization measurements (by way of electrometer charge accumulation studies), as well as high-resolution calorimetric measurements performed either in the a.c. or the relaxation mode¹³. The relaxation runs measure the total enthalpy changes (ΔH) including the latent heat (L), whereas the a.c. runs measure the continuous variations of the enthalpy¹³ (δH). Therefore by comparing results of two modes, not only the presence but also the precise value of the latent heat could be revealed¹³.

The temperature dependence of the excess heat capacity of PMN-PT with $x = 0.295$ (near the morphotropic boundary) is shown in Fig. 1a for different applied electric fields. The electric bias field was applied along the [111] direction. For low electric fields, there is a huge difference between a.c. and relaxation runs. This is due to the presence of latent heat, and demonstrates the existence of a first-order ferroelectric to paraelectric, that is, tetragonal to cubic, phase transition. With increasing electric field the latent heat diminishes. For $E = 1.5 \text{ kV cm}^{-1}$ the system is near critical ($L = 0 \text{ J g}^{-1}$), whereas for $E = 2 \text{ kV cm}^{-1}$ and $E = 4 \text{ kV cm}^{-1}$ a broad supercritical evolution can be observed.

It should be stressed that it is the supercritical behaviour that distinguishes the critical end point from a tri-critical point where a line of first-order transitions meets a line of second-order transitions. Whereas in the supercritical case the anomalies in the response functions become non-critical, smeared out, and finally disappear, there are sharp critical anomalies at the line of second-order transitions. Such a response, typical for the critical end point, was observed in solid ferroelectric NaNO_2 and various liquid crystal systems^{14–16}.

A small hysteresis of $\sim 2 \text{ K}$ was observed in the PMN-PT system between heating and cooling runs at typical cooling/heating rates of $0.5\text{--}1 \text{ K h}^{-1}$ and at low electric fields. The hysteresis gradually

¹Jozef Stefan Institute, Ljubljana, Slovenia. ²Institute of Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic.

diminished to zero as the field was increased beyond that of the critical end point value, E_{CL} . The total enthalpy ΔH exhibits a peak at the critical end point (Fig. 1b). A similar behaviour is seen in quasi-static dielectric measurements (Fig. 1c). Detailed experiments show that the critical end point exists for $x = 0.295$ at $E_{CL} = 1.3 \text{ kV cm}^{-1}$ (Fig. 1b inset).

In Fig. 2a we present the temperature dependence of the quasi-static polarization for PMN-PT with $x = 0.295$ at bias electric fields below and above the critical one. Four different anomalies have been

observed in the temperature dependencies of the dielectric constant, polarization, and most important, in the heat capacity even in zero electric field. This shows the presence of at least three new phases, between the tetragonal (T) and rhombohedral (R) ferroelectric states (Fig. 2b). Similarly to pure PMN, the polarization data demonstrate that the first-order relaxor to ferroelectric transition line ends in an isolated $E(T)$ critical end point with $E_{CL} = 1.3 \text{ kV cm}^{-1}$, above which the difference between the different phases disappears and a continuous supercritical evolution in various physical quantities can be observed. In pure PMN, on the other hand, the quasi-static polarization measurements (Fig. 2c) suggest a much higher critical field, $E_{CL} = 4 \text{ kV cm}^{-1}$.

According to our polarized light microscopy data, the polarization rotates under the applied [111] field at fixed temperature (or with

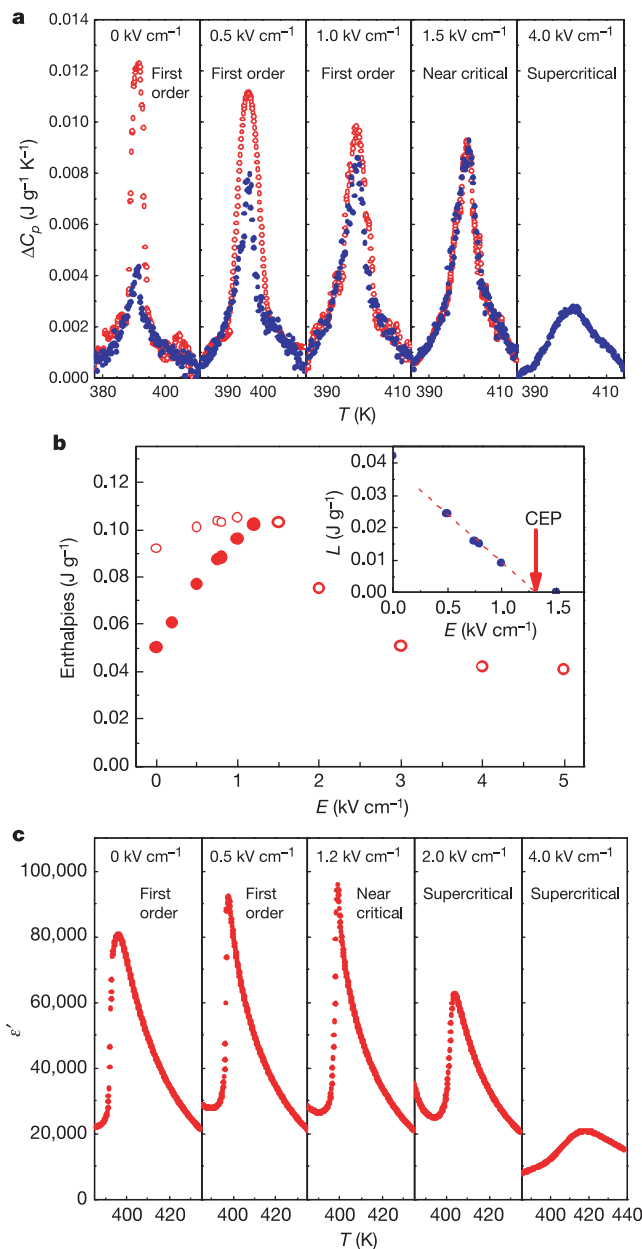


Figure 1 | Critical behaviour near the morphotropic phase boundary of PMN-PT. **a**, Temperature dependence of the excess heat capacity data obtained in PMN-PT $x = 0.295$ at various bias electric fields, E , shown at the top of each panel. Filled (blue) and open (red) circles represent data obtained in the a.c. and relaxation mode, respectively. The nature of the paraelectric to tetragonal phase transition is denoted by 'first order', 'near critical' and 'supercritical' labels. **b**, Variation of the total enthalpy ΔH (red open circles), the continuous part of the enthalpy δH (red filled circles), and inset, the latent heat L (blue filled circles), as a function of the electric field in PMN-PT $x = 0.295$. The position of the critical end point (CEP) is denoted by an arrow. **c**, Temperature dependence of the dielectric data obtained in PMN-PT $x = 0.295$ at various bias electric fields.

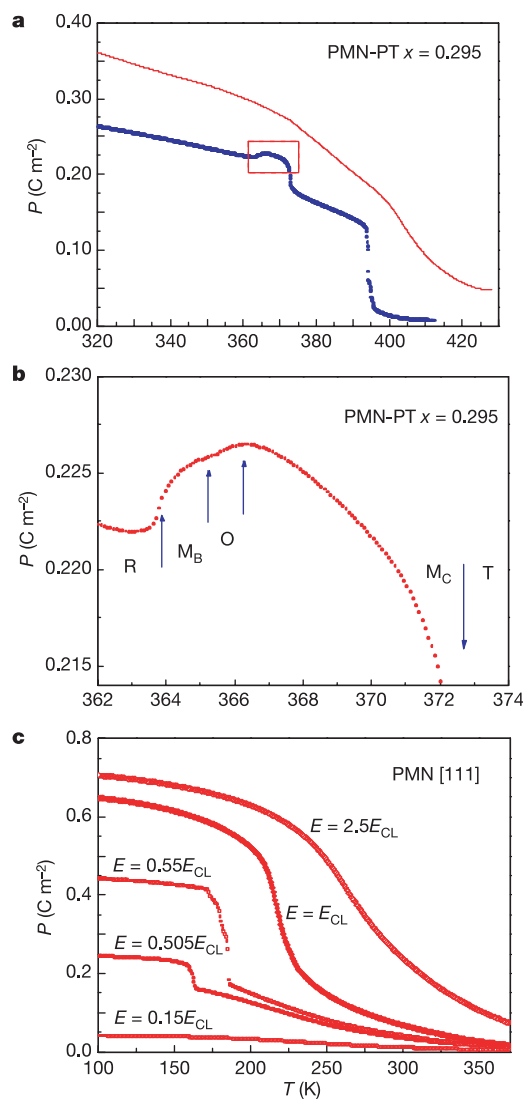


Figure 2 | Variation of the polarization order parameter as function of the electric field and temperature in PMN-PT and pure PMN single crystal.

a, Quasi-static polarization (P) data obtained in PMN-PT $x = 0.295$ along the [111] axis at $E = 0.2 \text{ kV cm}^{-1}$ (filled blue circles) and $E = 2.0 \text{ kV cm}^{-1}$ (solid red line), that is, at bias electric fields below and above the critical value of $E_{CL} = 1.3 \text{ kV cm}^{-1}$. Box encloses area shown in **b**. **b**, Close-up of additional anomalies observed in the polarization related to monoclinic (M_C), orthorhombic (O) and monoclinic (M_B) phases^{1,12} between the tetragonal (T) and rhombohedral (R) phase. **c**, Quasi-static polarization data obtained in pure PMN [111] single crystal ($x = 0$) at several different bias electric fields below and above the critical value of $E_{CL} \approx 4 \text{ kV cm}^{-1}$.

increasing temperature at zero field) from the tetragonal [001] direction into the (010) monoclinic (M_C) plane and then to the orthorhombic [101] direction. From there it goes to the rhombohedral [111] direction via the (101) monoclinic (M_B) plane. The sequence of the observed intermediate phases are thus $T-M_C-O-M_B-R$. This is in accordance with previous findings^{11,12} in $BaTiO_3$ and PMN-PT with $x = 0.32$.

It should be noted that recent work of Wu and Cohen¹⁷ predicts a morphotropic transition as a function of pressure (P), even in pure PT. Although in their theoretical T - P phase diagram they found two tricritical points and not critical end points, this work bolsters the idea of a critical line in the phase diagram that crosses the T - P plane at $x = 0$. They also show the important fact that the morphotropic phase boundaries and the giant piezoelectric effect do not require intrinsic disorder.

The E - T - x phase diagram of PMN-PT is shown in Fig. 3a. It shows that the plane of first-order ferroelectric-paraelectric phase transitions ends in a line of critical end points (LCEP), above which the system is supercritical. The line of critical end points shifts to lower electric field values when the PT concentration approaches the morphotropic rhombohedral-tetragonal phase boundary, thus explaining the enhancement of the piezoelectric response near this

boundary. The section of the phase diagram for $x = 0.295$, showing the critical end point, is presented in Fig. 3b.

The electric field dependence of the piezoelectric strain-coefficient d_{31} is shown in Fig. 4a for PMN-PT with $x = 0.295$. It exhibits a maximum at the electric field corresponding to the critical end point. As shown in Fig. 1b and c, both the total enthalpy and dielectric constant variations also exhibit a peak at the critical end point. Similar behaviour was also observed for other concentrations of PT.

Whereas the enthalpy changes for the two intermediate transitions, M_B-O and $O-M_C$, were always below 1 mJ g^{-1} , our heat capacity (Fig. 4b) and dielectric data show that the polarization rotation from the M_C to the T phase would require an enthalpy change of 35 mJ g^{-1} in the $E = 0$ plane, corresponding to a necessary field of $\sim 14 \text{ kV cm}^{-1}$ to induce this rotation. The necessary field decreases to 1.5 kV cm^{-1} at the critical end point, corresponding to an enthalpy change of only 4 mJ g^{-1} . Similarly, the enthalpy required for the $R-M_B$ polarization rotation (Fig. 4b inset) decreases on approaching the critical end point from 5 mJ g^{-1} (2 kV cm^{-1}) to below 1 mJ g^{-1} ($< 0.4 \text{ kV cm}^{-1}$). This demonstrates that for polarization rotations and the piezoelectric response the proximity of the critical end point is even more important than the proximity of the morphotropic phase boundary.

In the present work, we have found a line of critical end points of the liquid-vapour type in the E - T - x phase diagram of the PMN-PT system. The electric fields necessary for polarization rotations and the involved change of energy significantly decrease as the line of critical end points is approached, thus showing that the existence of the critical end point is at the heart of the giant electromechanical response of ferroelectric relaxors. This might be a central issue in the development and engineering of new advanced materials for electromechanical converters. Preliminary results (Z.K. and R.B.,

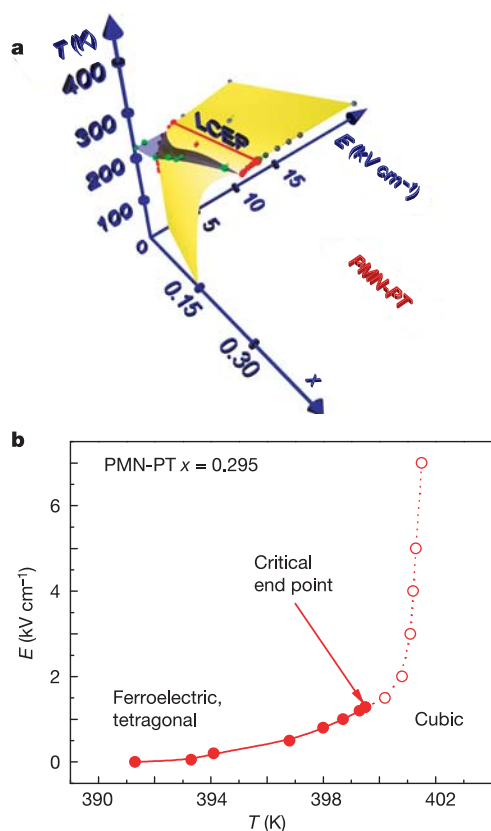


Figure 3 | The E - T - x phase diagram of the PMN-PT system.

a, Experimentally determined E - T - x phase diagram of the PMN-PT system. Circles of different colours represent experimentally determined position of phase transitions. Red line denotes the line of critical end points (LCEP). Red circles denote first-order ferroelectric phase transitions. Blue circles denote the region of the conversion plane where supercritical evolution was observed. Green circles denote the temperatures at which the longest dielectric relaxation time diverges—that is, the blue coloured plane together with the yellow coloured plane limit the low-temperature glassy region of the phase diagram. **b**, Section of the E - T - x phase diagram for $x = 0.295$. The open circles above the critical end point denote the positions of the smeared-out anomalies that gradually disappear at higher fields in the supercritical region. The labelling of the phases refers to the zero-electric-field phases.

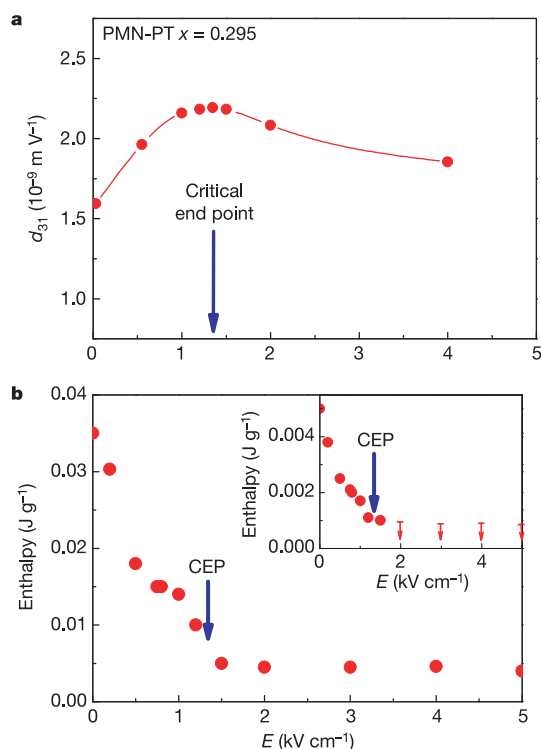


Figure 4 | Piezoelectric response and enthalpies related to the polarization rotations. **a**, Electric field dependence of the piezoelectric strain-coefficient d_{31} in the vicinity of the critical end point for PMN-PT $x = 0.295$. **b**, Electric field dependence of the enthalpy related to the M_C - T and $R-M_B$ (inset) polarization rotations. Red arrows in the inset denote the upper limit of the $R-M_B$ enthalpy.

unreported data) indicate that the above critical behaviour in the E dimension is not limited to the PMN-PT system, but also seems to exist in other relaxor materials, such as lead lanthanum zirconate titanate ceramics and strontium-barium niobate.

Received 13 December 2005; accepted 26 April 2006.

1. Cross, I. E. Relaxor ferroelectrics. *Ferroelectrics* **76**, 241–267 (1987).
2. Cross, I. E. in *Ferroelectric Ceramics* (eds Setter, N. & Colla, E. L.) 1–85 (Birkhauser, Berlin, 1993).
3. Samara, S. A. The relaxation properties of compositionally disordered ABO_3 perovskites. *J. Phys. Condens. Matter* **15**, R367–R411 (2003).
4. Uchino, K. *Piezoelectric Actuators and Ultrasonic Motors* (Kluwer Academic, Boston, 1996).
5. Colla, V. E., Yushin, N. K. & Viehland, D. Dielectric properties of $(\text{PMN})_{1-x}(\text{PT})_x$ single crystals for various electrical and thermal histories. *J. Appl. Phys.* **83**, 3298–3304 (1998).
6. Xu, G., Viehland, D., Li, J. F., Gehring, P. M. & Shirane, G. Evidence of decoupled lattice distortions and ferroelectric polarization in the relaxor system PMN-xPT. *Phys. Rev. B* **68**, 2410–2414 (2003).
7. Höchli, U. T., Knorr, K. & Loidl, A. Orientational glasses. *Adv. Phys.* **39**, 405–615 (1990).
8. Park, S. E. & Shrout, T. R. Ultrahigh strain and piezoelectric behaviour in relaxor based ferroelectric single crystal. *J. Appl. Phys.* **82**, 1804–1811 (2003).
9. Fu, H. & Cohen, R. E. Polarization rotation mechanism for ultrahigh electromechanical response in single crystal piezoelectric. *Nature* **403**, 281–283 (2000).
10. Li, J. Y., Rogan, R. C., Üstündag, E. & Bhattacharya, K. Domain switching in polycrystalline ferroelectric ceramics. *Nature Mater.* **41**, 776–781 (2005).
11. Davis, M., Damjanovic, D. & Setter, N. Electric-field-, temperature-, and stress-induced phase transitions in relaxor ferroelectric single crystals. *Phys. Rev. B* **73**, 014115 (2006).
12. Bai, F. et al. X-ray and neutron diffraction investigations of the structural phase transformation sequence under electric field in $0.7\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})-0.3\text{PbTiO}_3$ crystal. *J. Appl. Phys.* **96**, 1620–1627 (2004).
13. Yao, H., Ema, K. & Garland, C. W. Nonadiabatic scanning calorimeter. *Rev. Sci. Instrum.* **69**, 172–178 (1998).
14. Durand, D., Denoyer, F., Lefur, D., Currat, R. & Bernard, L. Neutron diffraction study of sodium nitrite in an applied electric field. *J. Phys. (Paris) Lett.* **44**, L207–L216 (1983).
15. Yao, H., Chan, T. & Garland, C. W. Smectic-C–smectic-I critical point in a liquid crystal mixture: Static and dynamic thermal behavior. *Phys. Rev. E* **51**, 4585–4597 (1995).
16. Kutnjak, Z. et al. Critical point for the blue-phase-III–Isotropic phase transition in chiral liquid crystals. *Phys. Rev. E* **53**, 4955–4963 (1996).
17. Wu, Z. & Cohen, R. E. Pressure-induced anomalous phase transitions and colossal enhancement of piezoelectricity in PbTiO_3 . *Phys. Rev. Lett.* **95**, 037601 (2005).

Acknowledgements This research was supported by the Slovenian Research Agency.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Z.K. (zdravko.kutnjak@ijs.si).

LETTERS

Broad-band optical parametric gain on a silicon photonic chip

Mark A. Foster¹, Amy C. Turner², Jay E. Sharping¹, Bradley S. Schmidt², Michal Lipson² & Alexander L. Gaeta¹

Developing an optical amplifier on silicon is essential for the success of silicon-on-insulator (SOI) photonic integrated circuits. Recently, optical gain with a 1-nm bandwidth was demonstrated using the Raman effect^{1–9}, which led to the demonstration of a Raman oscillator^{10,11}, lossless optical modulation¹² and optically tunable slow light¹³. A key strength of optical communications is the parallelism of information transfer and processing onto multiple wavelength channels. However, the relatively narrow Raman gain bandwidth only allows for amplification or generation of a single wavelength channel. If broad gain bandwidths were to be demonstrated on silicon, then an array of wavelength channels could be generated and processed, representing a critical advance for densely integrated photonic circuits. Here we demonstrate net on/off gain over a wavelength range of 28 nm through the optical process of phase-matched four-wave mixing in suitably designed SOI channel waveguides. We also demonstrate wavelength conversion in the range 1,511–1,591 nm with peak conversion efficiencies of +5.2 dB, which represents more than 20 times improvement on previous four-wave-mixing efficiencies in SOI waveguides^{14–17}. These advances allow for the implementation of dense wavelength division multiplexing in an all-silicon photonic integrated circuit. Additionally, all-optical delays¹⁸, all-optical switches¹⁹, optical signal regenerators²⁰ and optical sources for quantum information technology²¹, all demonstrated using four-wave mixing in silica fibres, can now be transferred to the SOI platform.

Amplification through four-wave mixing (FWM)²² is a nonlinear optical process derived from the third-order nonlinear susceptibility, $\chi^{(3)}$, of a material. In silicon, the lowest-order optical nonlinearity is third order, which gives rise to an intensity dependent refractive index $n = n_0 + n_2 I$, where n_0 is the linear refractive index, $n_2 = 12\pi^2 \chi^{(3)} / n_0 c$ is the nonlinear index coefficient, c is the speed of light, and I is the optical intensity. The measured n_2 values in silicon are in the range $(4–9) \times 10^{-14} \text{ cm}^2 \text{ W}^{-1}$ (refs 14, 23, 24), which is approximately 200 times that of silica glass. As shown in Fig. 1, in FWM two pump photons at frequency ω_{pump} are converted to signal and idler photons at respective frequencies ω_{signal} and ω_{idler} such that $2\omega_{\text{pump}} = \omega_{\text{signal}} + \omega_{\text{idler}}$, which can lead to amplification of the signal wave. In addition, the pump wave experiences self-phase modulation (SPM) and induces cross-phase modulation (XPM) on the signal and idler waves, which is twice as large as the pump SPM. The gain experienced by the signal wave depends on the phase-mismatch Δk between the propagation constants of the waves and on the nonlinear effects of SPM and XPM such that²²

$$\Delta k = 2\gamma P_{\text{pump}} - \Delta k_L \quad (1)$$

where $\Delta k_L = 2k_{\text{pump}} - k_{\text{signal}} - k_{\text{idler}}$ is the phase mismatch due to linear dispersion, k_{pump} , k_{signal} and k_{idler} are respectively the pump, signal and idler wavenumbers, P_{pump} is the pump power, $\gamma = \omega_{\text{pump}} n_2 / c A_{\text{eff}}$ is the effective nonlinearity of the waveguide, and

A_{eff} is the effective area of the propagating mode. The FWM gain coefficient g is given by²²

$$g = [\gamma P_{\text{pump}} \Delta k_L - (\Delta k_L/2)^2]^{\frac{1}{2}} \quad (2)$$

and neglecting the effects of pump depletion, the observed signal gain G_{signal} due to FWM is²²

$$G_{\text{signal}} = \frac{P_{\text{signal}}^{\text{out}}}{P_{\text{signal}}^{\text{in}}} = 1 + \left(\frac{\gamma P_{\text{pump}}}{g} \sinh(gL) \right)^2 \quad (3)$$

where $P_{\text{signal}}^{\text{out}}$ and $P_{\text{signal}}^{\text{in}}$ are the output and input signal powers, respectively, and L is the interaction length. The peak gain occurs when the phase shift due to SPM and XPM is compensated by the mismatch in the propagation constants of the pump, signal and idler waves. This requires the pump wave to have a larger propagation vector than the average of the signal and idler, which occurs when the group-velocity dispersion (GVD) is anomalous (that is, $d^2 k_{\text{pump}} / d\omega^2 < 0$). In optical fibres, FWM is routinely implemented in the telecommunications band beyond 1.3 μm , where the GVD of silica glass is anomalous²².

Four-wave mixing was previously observed in SOI waveguides with normal GVD, but the conversion occurred over a narrow band, the efficiencies were minimal, and net gain was not observed owing to the lack of phase-matching^{14–17}. At wavelengths near 1.5 μm , silicon exhibits very large normal GVD owing to the proximity of the absorption band edge at 1.1 μm . Waveguide confinement introduces anomalous GVD, which is used in silica glass fibres to create net anomalous-GVD regions²⁵ several octaves from any material resonance. However, it is not clear that this waveguide contribution in highly confining SOI waveguides can overcome the material dispersion so close to the absorption band and yield net anomalous

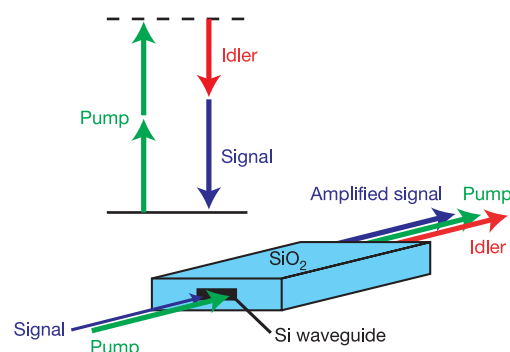


Figure 1 | Four-wave mixing with a degenerate pump. The four-wave-mixing process involves the conversion of two pump photons to a signal photon and an idler photon (top). By suitable design of the waveguide, momentum conservation (that is, phase-matching) is also satisfied and amplification of the idler occurs (bottom).

¹School of Applied and Engineering Physics, ²School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14853, USA.

GVD. We investigate this issue through simulations using a custom full-vector finite difference mode solver to calculate the GVD of the propagating mode. For accurate results, it is important to include the coupled effects of material and waveguide dispersion rather than treating them independently. For a small range of waveguide shapes and sizes, we predict^{26,27} and observe²⁷ the anomalous GVD required for phase-matching. In Fig. 2a, we plot the GVD for waveguides with different cross-sections, and the predicted FWM gain for a 2-cm propagation length and a 1-W peak pump power is shown in Fig. 2b. The waveguide with anomalous GVD ($300\text{ nm} \times 600\text{ nm}$) allows for phase-matching and is predicted to have a broad region of large amplification. In comparison, the smallest ($200\text{ nm} \times 400\text{ nm}$) and largest ($1.0\text{ }\mu\text{m} \times 1.5\text{ }\mu\text{m}$) waveguides have normal GVD and show minimal gain. The use of birefringence was suggested to achieve phase-matching in larger waveguides²⁸, but this additional complexity is not required for waveguides that exhibit anomalous GVD.

We experimentally investigate FWM in SOI waveguides with various cross-sectional geometries fabricated as previously

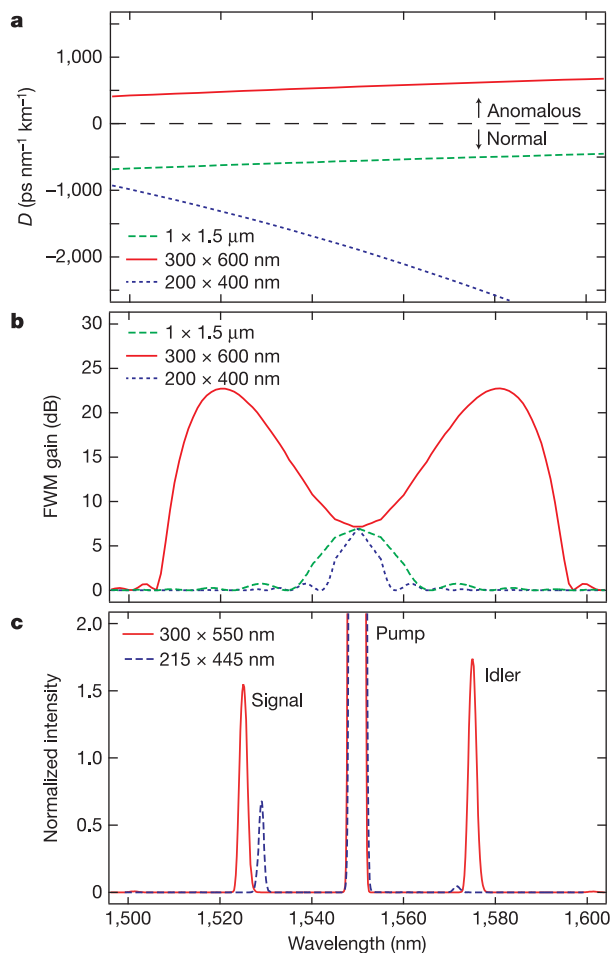


Figure 2 | Anomalous group-velocity dispersion and phase-matched four-wave mixing. **a**, Simulated group-velocity dispersion (GVD), denoted D , and **b**, four-wave mixing (FWM) gain for silicon waveguides with several cross-sectional dimensions. The FWM gain is calculated assuming a 1-W peak pump power and a 2-cm-long waveguide. The $300\text{ nm} \times 600\text{ nm}$ waveguide exhibits anomalous GVD ($D > 0$), allowing for phase-matching and large amplification. The smallest waveguide ($200\text{ nm} \times 400\text{ nm}$) and largest waveguide ($1\text{ }\mu\text{m} \times 1.5\text{ }\mu\text{m}$) have normal GVD ($D < 0$) and thus minimal gain. **c**, Experimentally observed FWM spectra from waveguides exhibiting anomalous GVD ($300\text{ nm} \times 550\text{ nm}$) and near-zero GVD ($215\text{ nm} \times 445\text{ nm}$). Both plots are normalized to the signal power with the pump off.

described²⁹ but with a $1\text{-}\mu\text{m}$ buried oxide layer. The pump and signal pulses are each derived from a femtosecond optical parametric oscillator centred at $1,550\text{ nm}$ with a 75-MHz repetition rate. The pump pulses are filtered to a 1-nm bandwidth centred at $1,550\text{ nm}$, and the signal pulses are filtered to a 1.5-nm bandwidth, tunable from $1,510\text{ nm}$ to $1,590\text{ nm}$. After filtering, the temporal durations of the pump and signal pulses are 3.5 ps and 2.4 ps , respectively. The pump pulses are amplified in an erbium-doped fibre amplifier and combined with the signal pulses using a wavelength division multiplexer. Both the pump and signal arms incorporate fibre polarization controllers to achieve TE polarization in the silicon waveguide. The light is coupled into the waveguide using a tapered lens fibre and an inverse taper mode converter on the silicon chip²⁹, which results in a

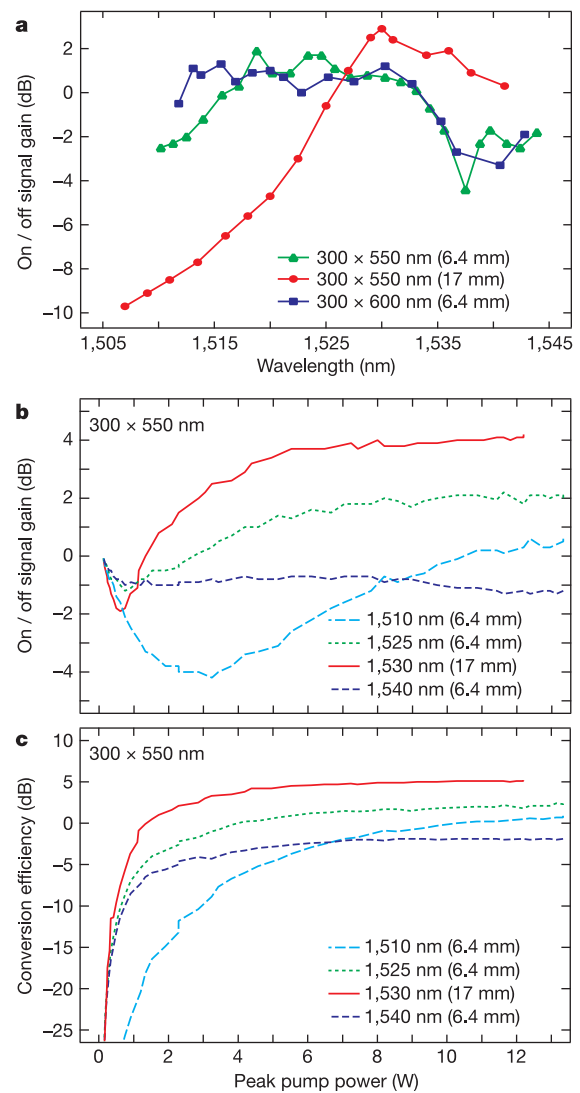


Figure 3 | Experimentally measured amplification. **a**, Measured on/off signal gain as a function of signal wavelength for several waveguide cross-sectional areas and lengths, which illustrates the gain sensitivity to waveguide dimensions. The key in **a** shows 'waveguide dimensions (length)', whereas the key in **b** and **c** shows 'signal wavelength (waveguide length)'. We observe a larger shift in the peak of the gain from the pump wavelength as the waveguide width is increased, which results from the reduced magnitude of anomalous GVD. **b**, **c**, Measured on/off signal gain (**b**) and idler conversion efficiency (**c**) as functions of peak pump power inside the waveguide for 6.4-mm - and 17-mm -long waveguides with $300\text{ nm} \times 550\text{ nm}$ cross-sections. Once a sufficient pump power is reached to achieve phase-matching at the observed wavelength, we measure a net nonlinear amplification of the signal.

measured coupling loss ranging from -12.3 dB to -13.3 dB. The large loss was due to poor overlap between the mode of the lens fibre and that of the inverse taper. The measured linear propagation loss in the waveguides ranges from -1.1 dB cm $^{-1}$ to -1.4 dB cm $^{-1}$ at $1,550$ nm. These values are determined by comparing the loss in waveguides of two lengths differing by 7.4 mm. These results were compared over six different waveguides for a consistent measurement. All the waveguides incorporated the same number of bends and taper couplers. The inverse taper ensures excitation of the fundamental mode for multi-mode waveguides. The light leaving the silicon waveguide is collected and sent to an optical spectrum analyser for detection.

Typically observed FWM signals for waveguides with anomalous GVD and near-zero GVD are shown in Fig. 2c. Using a 6.4 -mm-long 300 nm $\times$ 550 nm waveguide with a peak pump power of 11.0 W inside the waveguide, we observe net gain on the signal (1.9 dB) and highly efficient wavelength conversion ($+2.4$ dB). We expect that the idler wave carries more power than the signal owing to temporal and wavelength dependent effects of the nonlinear losses. For comparison, we plot the signal with near-zero GVD in an 8 -mm-long, 215 nm $\times$ 445 nm waveguide with a peak pump power of 8.0 W inside the waveguide. Despite the higher nonlinearity of this waveguide, FWM yields a net loss on the signal (-1.7 dB) and small conversion efficiency (-13.2 dB) owing to the absence of phase-matching.

The gain bandwidth for this system is illustrated in Fig. 3a, which shows the measured on/off signal gain as a function of signal wavelength for several waveguide cross-sections and lengths. We define the on/off signal gain and idler conversion efficiency as the signal and idler power, respectively, leaving the waveguide with the pump beam on divided by the signal power leaving the waveguide with the pump beam off^{6,17}. The effects of phase-matching are evident as the peak of the gain is shifted significantly from the pump wavelength. As expected, the magnitude of the shift increases for waveguides with decreasing GVD. Experimentally, we observe smaller gain than predicted by the simple model used for Fig. 2b owing to the combined effects of nonlinear absorption, free-carrier absorption and pump depletion. In Fig. 3b and c we plot the measured on/off signal gain (Fig. 3b) and idler conversion efficiency (Fig. 3c) as functions of peak pump power inside the waveguide for a 300 nm $\times$ 550 nm, 6.4 -mm-long waveguide and several signal wavelengths. From these curves, it is evident that the nonlinear loss mechanisms of two-photon absorption^{23,24} and free-carrier generation^{4,30} are dominant at low pump powers. With increased pump power, the FWM process becomes phase-matched at the wavelength of observation, and for wavelengths sufficiently far from the pump, this gain eventually exceeds the nonlinear losses, resulting in net nonlinear gain. For the 6.4 -mm-long waveguides, we observe a peak net on/off gain of 1.9 dB, and a net on/off gain can be produced from $1,512$ nm to $1,535$ nm with a pump laser tuned to $1,550$ nm (Fig. 3a).

A consequence of this FWM process is the ability to perform efficient wavelength conversion. As seen in Fig. 3c, we observe peak conversion efficiencies ranging from $+3$ dB to -2 dB for signal wavelengths from $1,511$ nm to $1,545$ nm, which corresponds to idler wavelengths from $1,555$ nm to $1,591$ nm. We were limited to this range owing to the bandwidth of the femtosecond source pulses. Larger wavelength conversion bandwidths are expected, as we observe conversion efficiency of $+3$ dB at the edge ($1,511$ nm) of the signal tuning range.

To offset nonlinear absorption and increase the gain, we implement a longer waveguide of 17 -mm length. This waveguide has a cross-section of 300 nm $\times$ 550 nm and four 100 - μ m-radius bends. The on/off signal gain is plotted as a function of wavelength in Fig. 3a. For a peak pump power of 4.2 W inside the waveguide, we see a peak on/off gain of 2.9 dB and an on/off gain bandwidth from $1,525$ nm to $1,540$ nm. Nonlinear absorption dominates outside the

phase-matching region and reaches -9.7 dB. In Fig. 3b and c, we plot the on/off signal gain and idler conversion efficiency as functions of pump power for this longer waveguide, and we observe a peak gain and efficiency of 4.2 dB and $+5.2$ dB, respectively. We expect that optimization of the waveguide length and GVD will yield further reductions in pump power requirements. Including linear propagation loss (-1.4 dB cm $^{-1}$), this 4.2 dB on/off gain yields a net gain of 1.8 dB on the signal after 1.7 cm of propagation. Observation of net gain allows for the possibility of parametric oscillation with the inclusion of on-chip resonating structures for feedback. Furthermore, comparing the largest observed gain to the observed loss outside the FWM bandwidth, we calculate the gain arising from FWM to be at least 13.9 dB.

We have demonstrated broadband phase-matched FWM optical amplification and frequency conversion in an all-silicon photonic chip. We expect that additional increases in performance can be achieved by implementing techniques to remove free carriers and minimize nonlinear losses^{9,11}. The ability to amplify and convert wavelengths over this wide range with a single pump laser, all within the communications band, dramatically enhances the signal processing capabilities of all-silicon photonic integrated circuits. Following these findings, we expect that numerous applications of FWM, already demonstrated in silica optical fibers, will be transferred to silicon photonic integrated circuits, including all-optical switching, optical signal regeneration and optical sources for quantum information technology.

Received 9 January; accepted 23 May 2006.

- Claps, R., Dimitropoulos, D., Raghunathan, V., Han, Y. & Jalali, B. Observation of stimulated Raman amplification in silicon waveguides. *Opt. Express* **11**, 1731–1739 (2003).
- Espinola, R. L., Dadap, J. I., Osgood, R. M. Jr, McNab, S. J. & Vlasov, Y. A. Raman amplification in ultrasmall silicon-on-insulator wire waveguides. *Opt. Express* **12**, 3713–3718 (2004).
- Liu, A., Rong, H., Paniccia, M., Cohen, O. & Hak, D. Net optical gain in a low loss silicon-on-insulator waveguide by stimulated Raman scattering. *Opt. Express* **12**, 4261–4268 (2004).
- Rong, H. et al. Raman gain and nonlinear optical absorption measurements in a low-loss silicon waveguide. *Appl. Phys. Lett.* **85**, 2196–2198 (2004).
- Xu, Q., Almeida, V. R. & Lipson, M. Time-resolved study of Raman gain in highly confined silicon-on-insulator waveguides. *Opt. Express* **12**, 4437–4442 (2004).
- Liang, T. K. & Tsang, H. K. Efficient Raman amplification in silicon-on-insulator waveguides. *Appl. Phys. Lett.* **85**, 3343–3345 (2004).
- Boyras, O. & Jalali, B. Demonstration of 11 dB fiber-to-fiber gain in a silicon Raman amplifier. *Electron. Express* **1**, 429–434 (2004).
- Xu, Q., Almeida, V. R. & Lipson, M. Demonstration of high Raman gain in a submicrometer-size silicon-on-insulator waveguide. *Opt. Lett.* **30**, 35–37 (2005).
- Jones, R. et al. Net continuous wave optical gain in a low loss silicon-on-insulator waveguide by stimulated Raman scattering. *Opt. Express* **13**, 519–525 (2005).
- Rong, H. et al. An all-silicon Raman laser. *Nature* **433**, 292–294 (2005).
- Rong, H. et al. A continuous-wave Raman silicon laser. *Nature* **433**, 725–728 (2005).
- Jones, R. et al. Lossless optical modulation in a silicon waveguide using stimulated Raman scattering. *Opt. Express* **13**, 1716–1723 (2005).
- Okawachi, Y. et al. All-optical slow-light on a photonic chip. *Opt. Express* **14**, 2317–2322 (2006).
- Fukuda, H. et al. Four-wave mixing in silicon wire waveguides. *Opt. Express* **13**, 4629–4637 (2005).
- Espinola, R. L., Dadap, J. I., Osgood, R. M. Jr, McNab, S. J. & Vlasov, Y. A. C-band wavelength conversion in silicon photonic wire waveguides. *Opt. Express* **13**, 4341–4349 (2005).
- Yamada, K. et al. All-optical efficient wavelength conversion using silicon photonic waveguide. *IEEE Photon. Technol. Lett.* **18**, 1046–1048 (2006).
- Rong, H., Kuo, Y. H., Liu, A., Paniccia, M. & Cohen, O. High efficiency wavelength conversion of 10 Gb/s data in silicon waveguides. *Opt. Express* **14**, 1182–1188 (2006).
- Sharping, J. E. et al. All-optical, wavelength and bandwidth preserving, pulse delay based on parametric wavelength conversion and dispersion. *Opt. Express* **13**, 7872–7877 (2005).
- Lin, Q. et al. 40-Gb/s optical switching and wavelength multicasting in a two-pump parametric device. *IEEE Photon. Technol. Lett.* **17**, 2376–2378 (2005).
- Ciaramella, E. & Trillo, S. All-optical reshaping via four-wave mixing in optical fibers. *IEEE Photon. Technol. Lett.* **12**, 849–851 (2000).

21. Li, X., Voss, P. L., Sharping, J. E. & Kumar, P. Optical-fiber source of polarization-entangled photons in the 1550 nm telecom band. *Phys. Rev. Lett.* **94**, 053601 (2005).
22. Hansryd, J., Andrekson, A., Westlund, M., Li, J. & Hedekvist, P. Fiber-based optical parametric amplifiers and their applications. *IEEE Select. Topics Quant. Electron.* **8**, 506–520 (2002).
23. Tsang, H. K. *et al.* Optical dispersion, two-photon absorption and self-phase modulation in silicon waveguides at 1.5 μm wavelength. *Appl. Phys. Lett.* **80**, 416–418 (2002).
24. Dinu, M., Quochi, F. & Garcia, H. Third-order nonlinearities in silicon at telecom wavelengths. *Appl. Phys. Lett.* **82**, 2954–2956 (2003).
25. Foster, M. A., Moll, K. D. & Gaeta, A. L. Optimal waveguide dimensions for nonlinear interactions. *Opt. Express* **12**, 2880–2887 (2004).
26. Raghunathan, V., Claps, R., Dimitropoulos, D. & Jalali, B. Parametric Raman wavelength conversion in scaled silicon waveguides. *J. Lightwave Technol.* **23**, 2094–2102 (2005).
27. Turner, A. C. *et al.* Tailored anomalous-group velocity dispersion in silicon channel waveguides. *Opt. Express* **14**, 4357–4362 (2006).
28. Dimitropoulos, D., Raghunathan, V., Claps, R. & Jalali, B. Phase-matching and nonlinear optical processes in silicon waveguides. *Opt. Express* **12**, 149–160 (2004).
29. Almeida, V. R., Panepucci, R. R. & Lipson, M. Nanotapers for compact mode conversion. *Opt. Lett.* **28**, 1302–1304 (2003).
30. Liang, T. K. & Tsang, H. K. Role of free carriers from two-photon absorption in Raman amplification in silicon-on-insulator waveguides. *Appl. Phys. Lett.* **84**, 2745–2747 (2004).

Acknowledgements We acknowledge discussions with Y. Okawachi. This work was supported by the Center for Nanoscale Systems, supported by the NSF and the New York State Office of Science, Technology & Academic Research. M.A.F., J.E.S. and A.L.G. acknowledge support from the DARPA DSO Slow-Light Program.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.L.G. (a.gaeta@cornell.edu).

LETTERS

The Southern Ocean biogeochemical divide

I. Marinov^{1†}, A. Gnanadesikan², J. R. Toggweiler² & J. L. Sarmiento¹

Modelling studies have demonstrated that the nutrient and carbon cycles in the Southern Ocean play a central role in setting the air–sea balance of CO₂ and global biological production^{1–8}. Box model studies^{1–4} first pointed out that an increase in nutrient utilization in the high latitudes results in a strong decrease in the atmospheric carbon dioxide partial pressure (p_{CO_2}). This early research led to two important ideas: high latitude regions are more important in determining atmospheric p_{CO_2} than low latitudes, despite their much smaller area, and nutrient utilization and atmospheric p_{CO_2} are tightly linked. Subsequent general circulation model simulations show that the Southern Ocean is the most important high latitude region in controlling pre-industrial atmospheric CO₂ because it serves as a lid to a larger volume of the deep ocean^{5,6}. Other studies point out the crucial role of the Southern Ocean in the uptake and storage of anthropogenic carbon dioxide⁷ and in controlling global biological production⁸. Here we probe the system to determine whether certain regions of the Southern Ocean are more critical than others for air–sea CO₂ balance and the biological export production, by increasing surface nutrient drawdown in an ocean general circulation model. We demonstrate that atmospheric CO₂ and global biological export production are controlled by different regions of the Southern Ocean. The air–sea balance of carbon dioxide is controlled mainly by the biological pump and circulation in the Antarctic deep-water formation region, whereas global export production is controlled mainly by the biological pump and circulation in the Subantarctic intermediate and mode water formation region. The existence of this biogeochemical divide separating the Antarctic from the Subantarctic suggests that it may be possible for climate change or human intervention to modify one of these without greatly altering the other.

Recent palaeoceanographic literature points out that different parts of the Southern Ocean have responded differently to climate change^{9–11}. In an attempt to elucidate the complex mechanisms governing Southern Ocean carbon sequestration, we examine: first, which areas of the Southern Ocean control atmospheric p_{CO_2} levels; second, which areas control global biological export production; and third, whether there is a link between atmospheric p_{CO_2} and global export production.

A simple and effective probe for examining these questions is surface nutrient depletion, that is, forcing surface nutrients towards zero by increasing nutrient uptake and converting nutrients to export production. The effect is to reduce p_{CO_2} in surface water, driving CO₂ from the atmosphere into the ocean. Our approach follows previous studies⁵ and gives an upper limit estimate of the oceanic uptake that might result from complete removal of surface nutrients in a specified ocean region. This theoretical approach ignores complex phytoplankton growth limitations that are likely to limit nutrient depletion, such as light supply, iron, zooplankton grazing, and particle aggregation.

We deplete nutrients over nine Southern Ocean regions in the Princeton general circulation model (GCM; Table 1, Methods). The areas of interest are defined as surface waters with densities greater than or lighter than reference potential densities σ_θ of 27.6, 27.4, 27.3 and 27.1 kg m⁻³ (Fig. 1b). Table 1 summarizes the results of the experiments.

We find that atmospheric p_{CO_2} depends strongly on the location where nutrient depletion occurs. If we define a depletion efficiency as the ratio between the atmospheric p_{CO_2} drawdown and the surface area depleted in nutrients, Table 1 shows that nutrient depletion is more efficient at drawing-down atmospheric CO₂ in the southernmost regions (depletion experiments (Exps) S27.6, S27.4, S27.3, S27.1) than in the northern regions (Exps N27.6, N27.4, N27.3, N27.1). This suggests the existence of a Southern Ocean biogeochemical divide, south of which nutrient depletion has most impact on air–sea carbon exchange. How can we explain this finding?

The first order explanation lies with the existing large scale oceanic circulation. As a consequence of wind driven forcing, deep water rich in nutrients and dissolved inorganic carbon (DIC) upwells as Circumpolar Deep Water (CDW) to the south of the Antarctic Polar Front^{12–14}. It has long been known from observations^{14–16} that some portion of this water moves northwards in the surface Ekman layer and is incorporated into Antarctic Intermediate Water (AAIW) and Subantarctic Mode Water (SAMW), while some portion moves south and is subducted as Antarctic Bottom Water (AABW).

Figure 1a taken from ref. 10 summarizes this view of oceanic circulation (see also Methods). In this simple diagram, the lower (blue) circulation comprises the southward moving water subducted as AABW. The upper (red) circulation comprises the upwelling water

Table 1 | Southern Ocean nutrient depletion in the Princeton GCM

Depletion experiments	Region depleted	Area (% total)	Δ Atm. p_{CO_2} (p.p.m.)	Δ Prodn (GtC yr ⁻¹)	Efficiency of depletion (10 ⁻⁶ p.p.m. km ⁻²)
S30S	South of 30°S	29.32	-70.33	0.49	0.66
Antarctic experiments					
S27.1	$\sigma_\theta \geq 27.1$	12.19	-61.29	-1.09	1.4
S27.3	$\sigma_\theta \geq 27.3$	5.9	-47.62	-0.41	2.24
S27.4	$\sigma_\theta \geq 27.4$	3.65	-39.23	-0.06	2.98
S27.6	$\sigma_\theta \geq 27.6$	1.1	-23.10	0.63	5.85
Subantarctic experiments					
N27.1	$\sigma_\theta < 27.1$	17.13	-18.77	2.01	0.3
N27.3	$\sigma_\theta < 27.3$	23.42	-37.03	0.21	0.44
N27.4	$\sigma_\theta < 27.4$	25.66	-43.59	-0.08	0.47
N27.6	$\sigma_\theta < 27.6$	28.22	-53.50	-0.18	0.53

Column 2 defines the region depleted relative to the surface potential density σ_θ . Column 3 is the area depleted as a fraction of the total ocean area. Columns 4–6 show respectively changes in atmospheric p_{CO_2} , total global export production and efficiency of depletion (defined as the ratio of atmospheric p_{CO_2} drawdown and surface area depleted, in absolute value) after each depletion simulation. Antarctic nutrient depletion is more efficient than Subantarctic depletion. Atmospheric p_{CO_2} is calculated relative to an undepleted (control) atmospheric p_{CO_2} of 298.1 p.p.m. Undepleted global production is 9.40 GtC yr⁻¹. Results are shown at equilibrium for our regular gas exchange (standard) model. See Methods for comments on the Weddell Sea experiment (S27.6).

¹Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey 08540, USA. ²NOAA/Geophysical Fluid Dynamics Laboratory, PO Box 308, Forrestal Campus, Princeton, New Jersey 08542, USA. [†]Present address: Program in Atmospheres, Oceans, and Climate, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

which moves northward to the rest of the ocean at intermediate depths and returns to the Southern Ocean via the North Atlantic Deep Water (NADW)¹³. The surface separation between the two circulations is somewhere south of the Antarctic Polar Front, and corresponds to the divide between the Antarctic (deep water formation) region to the south from the Subantarctic (intermediate and mode water formation) region to the north.

We propose that this physical separation leads to significantly different responses to nutrient depletion in the two regions—and hence to the biogeochemical divide detected in our experiments. The key to understanding this result lies in the concept of biological pump efficiency.

The biological carbon pump is primarily produced by the photosynthetic formation of organic matter in the surface ocean and its subsequent transport to depth, where it is remineralized. ('Remineralized nutrients' are nutrients added to the ocean interior through remineralization; 'preformed nutrients' are biologically unutilized surface nutrients subducted into the ocean interior via newly formed deep water.) The biological pump acts to decrease surface preformed

nutrients, converting them to export production and increasing the concentration of remineralized nutrients deeper in the water column. Because of the association of nutrients with carbon in both photosynthesis and remineralization, the more efficient the conversion of preformed to remineralized nutrients, the more efficient the biological pump in sequestering carbon in the deep ocean and reducing atmospheric $p\text{CO}_2$, as recently discussed¹⁷.

Increased surface nutrient depletion in deep water formation regions lowers the concentration of preformed nutrients in newly formed deep waters. Since total PO_4^{3-} is constant in our experiments, a decrease in globally averaged preformed nutrients signals a more efficient conversion to remineralized nutrients, a more efficient biological pump and a stronger decrease in atmospheric $p\text{CO}_2$ (Fig. 2a and refs 1–3). The larger the decrease in preformed nutrients, the more efficient the biological pump and the oceanic carbon sequestration.

The most relevant preformed nutrients for the biological pump efficiency and for atmospheric $p\text{CO}_2$ are those present in deep water formation areas at the time deep water forms. Concentrations of preformed nutrients in high latitude, deep water formation regions are high, owing to inefficient biology. These nutrients are advected and mixed into the deep ocean; the preformed nutrient concentration relevant for the global biological carbon pump is therefore a linear combination of the surface preformed concentration in each deep water formation area weighted by the corresponding net contribution of the deep water source to the deep reservoir.

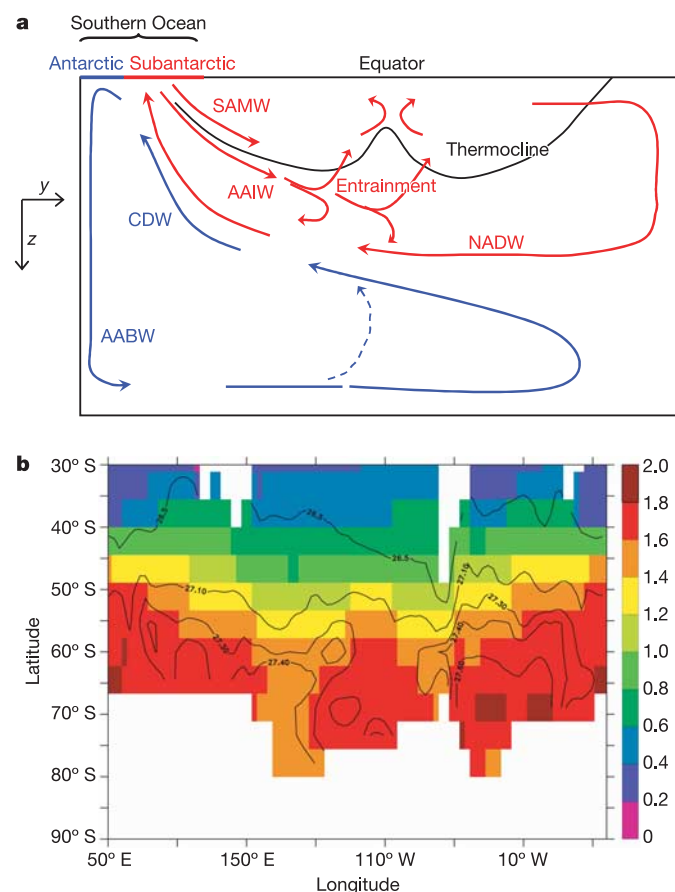


Figure 1 | Global circulation diagram relevant for the carbon cycle, and experimental set-up. a, Circumpolar Deep Water (CDW) upwelling to the Southern Ocean surface is either advected northward as part of the upper circulation (in red), or moves southward as lower circulation (in blue); this panel is adapted from ref. 10. The Antarctic–Subantarctic surface separation at the Antarctic Polar Front roughly corresponds to the $\sigma_o = 27.3$ outcrop in the Princeton GCM. See Methods for the simplifications used in this diagram. AABW, Antarctic Bottom Water; AAIW, Antarctic Intermediate Water; NADW, North Atlantic Deep Water; SAMW, Subantarctic Mode Water. **b**, Undepleted (control) annual mean surface PO_4^{3-} ($\mu\text{mol kg}^{-1}$) shown in colours. Geographical extent of the August σ_o surface outcrops shown as contours in units of kg m^{-3} . Surface nutrients are depleted everywhere south of these density outcrops (Exps S27.1, S27.3, S27.4, S27.6) and between these outcrops and 30°S (Exps N27.1, N27.3, N27.4, N27.6). σ_o is defined as $(\rho_{\text{pot}} - 1,000) + 0.025 (\text{kg m}^{-3})$, where ρ_{pot} is the actual potential density referenced to the surface.

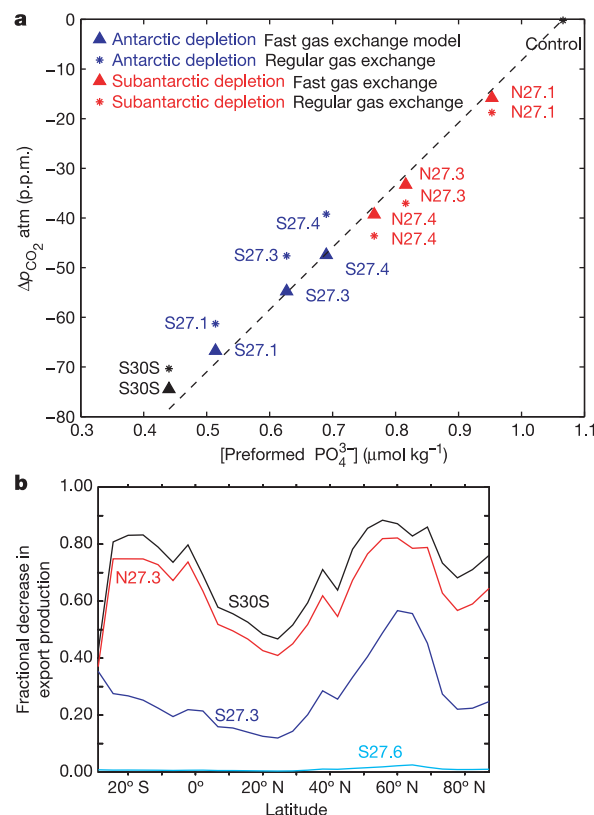


Figure 2 | Oceanic mechanisms for modifying atmospheric CO_2 and global production in the Princeton GCM. a, A decrease in globally averaged preformed PO_4^{3-} results in decreased atmospheric $p\text{CO}_2$. For all simulations, CO_2 drawdown is similar in our regular gas exchange (stars) and fast gas exchange (triangles) models. Antarctic nutrient depletions (blue triangles and stars) generally result in a larger decrease in preformed nutrients and stronger atmospheric $p\text{CO}_2$ drawdown than Subantarctic depletions (red triangles and stars). Undepleted/control case also shown (black); see Methods for details. **b**, Fractional decrease in export production following nutrient depletion experiments. Subantarctic depletion (Exp. N27.3) has more impact on global production than Antarctic depletion (Exp. S27.3).

Let us consider the main water masses in Fig. 1a: AABW, NADW, AAIW/SAMW. The balance of preformed nutrients between these sources determines to first order how much CO_2 the deep ocean holds. The surface preformed nutrient concentration is largest in AABW and smallest in the NADW formation region. The AABW ventilates the largest volume of deep water, followed by NADW, AAIW and SAMW. Because of higher surface preformed nutrient concentration and stronger deep ocean ventilation via AABW, Antarctic nutrient depletion has more impact on deep ocean preformed nutrients and results in larger atmospheric CO_2 drawdown compared to the Subantarctic depletion (compare Exps S27.3 and N27.3 in Figs 2a, 3). Interestingly, Antarctic convective regions are very efficient at sequestering atmospheric CO_2 (see Exp. S27.6 in Table 1 and Methods).

We now examine how nutrient depletion affects globally integrated biological export production. Southern Ocean nutrient depletion has two separate effects on export production^{6,18}. Locally, it increases production. Moreover, as the negative nutrient perturbation is advected by the upper circulation, subducted north of the Antarctic Polar Front as intermediate and mode waters and eventually mixed through the thermocline to reach both low latitude surface and high northern latitudes, both nutrients and export production decrease everywhere north of 30°S . The associated fractional decrease in export production outside the Southern Ocean can reach 75% (Fig. 2b), as also discussed in ref. 8.

Because the upper circulation is the primary mechanism by which nutrients are advected into low latitudes, depleting nutrients in

Subantarctic waters (Exp. N27.3) results in much larger decrease in export production north of 30°S than Antarctic depletion (Exp. S27.3). Note that observations of silica⁸ (defined as $\text{Si}(\text{OH})_4 - \text{NO}_3^-$) and radiocarbon¹⁹ also suggest a large role for Subantarctic nutrients in low latitude production. Figure 2b shows that potential density $\sigma_\theta = 27.3$ (kg m^{-3}) is roughly the Southern Ocean biogeochemical divide in our model, with areas south (north) of this boundary in the Antarctic (Subantarctic) region. Despite its considerable carbon uptake, the Weddell Sea is completely disconnected in terms of nutrient transport from the low latitudes; local nutrient changes do not affect nutrients and production north of 30°S (Exp. S27.6, Fig. 2b).

The decrease in atmospheric p_{CO_2} relative to the change in production south of 30°S (Fig. 3a) and north of 30°S (Fig. 3b) further highlights the differential response to depletion of the Antarctic and Subantarctic. Interestingly, there is no meaningful relationship between atmospheric p_{CO_2} and global export production (Fig. 3c). The results of Exps S27.3, S27.4 and S27.6 lie on the same p_{CO_2} versus production branch in Fig. 3a and b, confirming that in our model all points south of $\sigma_\theta = 27.3$ are in the Antarctic and respond similarly to nutrient depletion. For a given increase (decrease) in production south (north) of 30°S , the Antarctic is much better at taking up atmospheric CO_2 than the Subantarctic. The two-branch structure in Fig. 3a and b confirms the existence of two distinct biogeochemical regimes.

The relative contributions of different deep water types to the Antarctic and Subantarctic surface waters, an issue at present debated in the research community (see Methods), might affect the fraction of low latitude production driven by Southern Ocean nutrients (Fig. 2b). However, the existence of the biogeochemical divide is a robust result, which holds across GCMs with different Southern Ocean circulations and box models. Simple box model experiments (see Supplementary Information) show that slow Antarctic–Subantarctic surface mixing and the separation into northward and southward flows at the surface (rather than the separation of the lower and upper circulations at depth), together with the confinement of the AABW formation to the Antarctic, lead to the existence of the biogeochemical divide.

We acknowledge limitations to our model such as coarse resolution, inability to resolve shelf or slope processes and lack of a dynamic ice model. Although the exact sensitivity of atmospheric CO_2 to the global inventory of preformed nutrients (the regression slope in Fig. 2a) may vary among models and between our model and nature²⁰, the efficiency of the biological pump in drawing down atmospheric CO_2 is robustly given, to first order, by the deep ocean inventory of preformed nutrients. In turn, the preformed nutrients inventory depends on both surface unused nutrient concentration and deep ocean ventilation.

Our analysis shows clearly in the context of a realistic GCM that the physical separation between lower and upper oceanic circulations at the ocean surface (Fig. 1a) has a simple translation into a biogeochemical divide. The air–sea carbon balance is primarily determined by the amount of deep water formation and convection in the Antarctic, the unutilized nutrient concentration in this region and the area through which mixing with the deep waters occurs. By contrast, global and low latitude export production are set primarily by the amount of intermediate and deep water formation in the Subantarctic, and by the area and the unutilized nutrient concentration of the Subantarctic region. This suggests that it might be possible for climate change to modify atmospheric CO_2 without greatly modifying low latitude and global export production, and vice versa. We suggest that mechanisms that attempt to explain the lower atmospheric CO_2 during the Last Glacial Maximum—such as changes in deep water formation^{10,21} and oceanic stratification⁹, Fe fertilization²², and reduced ventilation due to increased ice cover²³—should take into account differences in carbon sequestration efficiency between the Antarctic and Subantarctic.

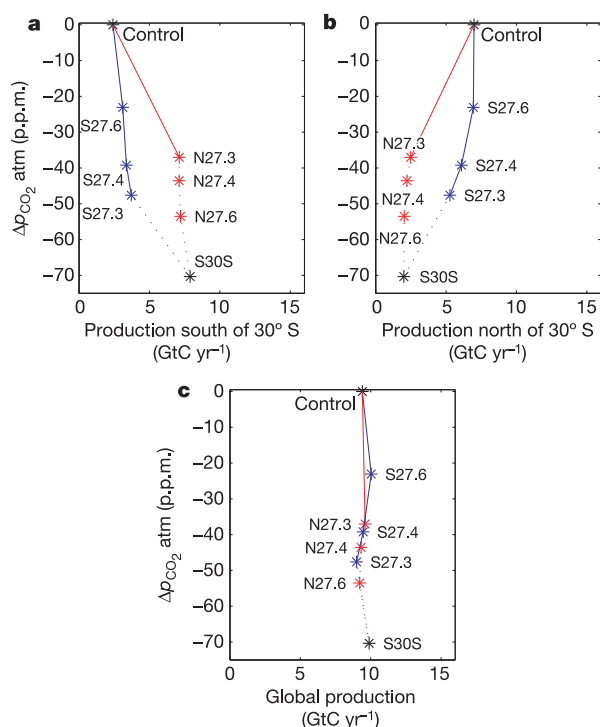


Figure 3 | Change in atmospheric p_{CO_2} versus export production after surface nutrient depletion simulations in the regular gas exchange model. Export production is integrated **a**, south of 30°S ; **b**, north of 30°S ; and **c**, globally. Undepleted/control scenario is shown in black, the Subantarctic branch in red, and the Antarctic branch in blue. A comparison of the two branches in **a** and **b** shows that Antarctic areas take up atmospheric CO_2 more efficiently (Exps S27.3, S27.4, S27.6), while areas primarily in the Subantarctic have a larger impact on global biological production (Exps N27.3, N27.4, N27.6). The rough Antarctic–Subantarctic boundary is $\sigma_\theta = 27.3$. Dotted lines connect scenarios in which nutrients are partially depleted in both Antarctic and Subantarctic areas. There is no robust relationship between global biological production and atmospheric p_{CO_2} .

METHODS

Model set-up. The GCM used in this study is the Geophysical Fluid Dynamics Laboratory Modular Ocean Model version 3 (MOM3) coupled with an OCMIP 2 biogeochemistry model²⁴. A single well-mixed atmospheric box is coupled to the ocean carbon cycle; the total carbon in the atmosphere and ocean is conserved. For a detailed description of our seasonally forced model (KVLW + AILW), which uses the Gent-McWilliams parameterization of subgrid scale processes, see ref. 25. New production in the undepleted case is obtained by forcing PO_4^{3-} towards observed seasonal Levitus values, $\text{PO}_4^{3-}\text{obs}$, in the upper 75 m of the model at each time step t , when $\text{PO}_4^{3-} > \text{PO}_4^{3-}\text{obs}$:

$$J_{\text{prod}}(x, y, z, t) = (\text{PO}_4^{3-}(x, y, z, t) - \text{PO}_4^{3-}\text{obs}(x, y, z, t))/T, \quad z < 75 \text{ m}$$

where $T = 30$ days, and x , y , and z are model longitude, latitude and depth, respectively. In the nutrient depletion cases we replace $\text{PO}_4^{3-}\text{obs}$ with zero in the above equation. Nutrients in each of the nine simulations are continuously depleted for a few thousand years until a new equilibrium has been achieved; globally integrated nutrients are kept constant during these experiments. Shown are pre-industrial results at equilibrium.

The simulations discussed here are performed with a soft tissue pump model, in which there is no calcium carbonate cycling, and surface temperature and salinity are set to 10°C and 34.7 p.s.u. (practical salinity units), for the purpose of the gas exchange calculations. We show elsewhere⁶ that these simplifications have a minimal impact on the results. Indeed, this model accounts for most of the atmospheric $p\text{CO}_2$ drawdown following Southern Ocean nutrient depletion in a model that includes the full carbonate and solubility components of the carbon pump. The model includes climatological sea-ice cover at monthly resolution as in OCMIP2²⁴.

The Weddell Sea in our model. The Antarctic region south of $\sigma_\theta = 27.6$ corresponds in our model to the Weddell Sea, a deep water formation region with high preformed nutrients and deep winter convection. Even though the Weddell Sea accounts for only 3.3% of the Southern Ocean, nutrient depletion in this area accounts for 30% of the 70 p.p.m. uptake following depletion of the entire Southern Ocean, suggesting that nutrient depletion in deep convective regions is particularly effective at sequestering CO_2 .

Simplification of complex overturning circulation. Figure 1a is a simplified two-dimensional diagram of a complex overturning circulation. The composition of the water upwelling in the Southern Ocean, and of the resulting waters moving northward and southward, are a matter of great debate in the oceanographic community. Numerical models such as ours^{25,26}, observations of silica⁴⁸ and radiocarbon¹⁹ and older observational syntheses¹⁶ suggest that a considerable amount of deep water (in particular of the NADW type) is transformed into lighter water in the Southern Ocean and entrained northward into the AAIW and SAMW. By contrast, recent data-based estimates²⁷ and inverse model studies²⁸ suggest that the Southern Ocean transformation might be quite small. These studies also suggest that a percentage of NADW is converted into surface waters south of the Antarctic Polar Front area where Upper Circumpolar Deep Water upwells, such that the lower and upper circulations in Fig. 1a are interconnected rather than separate. Box model studies (Supplementary Information) tend to suggest that this would have little impact on the biogeochemical divide.

Preformed nutrients and atmospheric $p\text{CO}_2$. Nutrient depletion simulations were separately run in two models: one with regular gas exchange and one with fast gas exchange (where oceanic CO_2 at each time step is set equal to atmospheric CO_2 calculated at the previous time step), as shown in Fig. 2a. For both models, a larger decrease in preformed nutrients signals a stronger biological uptake of atmospheric CO_2 and a more efficient biological pump^{1–3,17}.

If gas exchange is infinitely fast, surface DIC is in equilibrium with the atmosphere and the surface to deep DIC difference and atmospheric $p\text{CO}_2$ levels are entirely due to increased remineralization. Global mean preformed nutrients are the difference between total globally averaged nutrients (a constant in all our experiments) and global mean remineralized nutrients. Global mean remineralized nutrients are calculated as the surface to deep DIC difference in the fast gas exchange model divided by the Redfield ratio, $r_{\text{C:P}}$. Since the time step of the model limits how fast gas exchange is, there is a maximum error of 5–10% associated with this method.

While atmospheric $p\text{CO}_2$ is linearly related to preformed PO_4^{3-} in the fast gas exchange model, results for the regular gas exchange model are slightly off the fast gas exchange line. The deviation of these results from the fast gas exchange line is due to surface DIC disequilibrium (relative to the atmosphere) at the time of subduction. The small deviation of the regular model from the fast gas exchange line shows that in this particular GCM the oceanic CO_2 uptake is primarily driven by changes in remineralized nutrients. Changes in carbon uptake due to changes in surface DIC disequilibrium with nutrient depletion are relatively small.

Received 4 May 2005; accepted 16 May 2006.

- Sarmiento, J. L. & Toggweiler, J. R. A new model for the role of the oceans in determining atmospheric $p\text{CO}_2$. *Nature* **308**, 620–624 (1984).
- Knox, F. & McElroy, M. B. Changes in atmospheric CO_2 : Influence of the marine biota at high latitude. *J. Geophys. Res.* **89**, 4629–4637 (1984).
- Siegenthaler, U. & Wenk, T. H. Rapid atmospheric CO_2 variations and ocean circulation. *Nature* **308**, 624–626 (1984).
- Joos, F., Sarmiento, J. L. & Siegenthaler, U. Estimates of the effect of Southern Ocean iron fertilization on atmospheric CO_2 concentrations. *Nature* **349**, 772–775 (1991).
- Sarmiento, J. L. & Orr, J. C. Three-dimensional ocean model simulations of the impact of Southern Ocean nutrient depletion on atmospheric CO_2 and ocean chemistry. *Limnol. Oceanogr.* **36**, 1928–1950 (1991).
- Marinov, I. *Controls on the Air-Sea Balance of Carbon Dioxide*. PhD thesis, Princeton Univ. (2005).
- Caldeira, K. & Duffy, P. B. The role of the Southern Ocean in uptake and storage of anthropogenic carbon dioxide. *Science* **287**, 620–622 (2000).
- Sarmiento, J. L., Gruber, N., Brzezinski, M. A. & Dunne, J. P. High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* **427**, 56–60 (2004).
- Sigman, D. M. & Boyle, E. A. Glacial/interglacial variations in atmospheric carbon dioxide. *Nature* **407**, 859–869 (2000).
- Toggweiler, J. R., Russell, J. L. & Carson, S. R. Mid-latitude westerlies, atmospheric CO_2 and climate change during the Ice Ages. *Paleoceanography* **21**, PA2005, doi:10.1029/2005PA001154 (2006).
- Kohfeld, K. E., Le Quere, C., Harrison, S. P. & Anderson, R. F. Role of marine biology in glacial-interglacial CO_2 cycles. *Science* **308**, 74–78 (2005).
- Gordon, A. in *Antarctic Oceanography I* (ed. Reid, J. L.) Antarctic Res. Ser. Vol. 15, 169–203 (American Geophysical Union, Washington DC, 1971).
- Toggweiler, J. R. & Samuels, B. Effect of Drake Passage on the global thermohaline circulation. *Deep-sea Res.* **42**, 477–500 (1995).
- Wyrtki, K. The thermohaline circulation in relation to the general circulation in the oceans. *Deep-Sea Res.* **8**, 39–64 (1961).
- Sverdrup, H. U., Johnson, M. W. & Fleming, R. H. *The Oceans, Their Physics, Chemistry, and General Biology* (Prentice-Hall, New York, 1942).
- Schmitz, W. J. On the interbasin-scale thermohaline circulation. *Rev. Geophys.* **33**, 151–173 (1995).
- Ito, T. & Follows, M. J. Preformed phosphate, soft tissue pump and atmospheric CO_2 . *J. Mar. Res.* **63**, 813–839, doi:10.1357/0022240054663231 (2005).
- Dutkiewicz, S., Follows, M. J. & Parekh, P. Interactions of the iron and phosphorus cycles: a three dimensional model study. *Glob. Biogeochem. Cycles* **19**, GB1021, doi:10.1029/2004GB002342 (2005).
- Toggweiler, J. R., Dixon, K. & Broecker, W. S. The Peru upwelling and the ventilation of the South Pacific thermocline. *J. Geophys. Res.* **96**, 20467–20497 (1991).
- Toggweiler, J. R., Murnane, R., Carson, S., Gnanadesikan, A. & Sarmiento, J. L. Representation of the carbon cycle in box models and GCMs—Part 2. Organic pump. *Glob. Biogeochem. Cycles* **17**, 1027, doi:10.1029/2001GB001841 (2003).
- Toggweiler, J. R. Variation of atmospheric CO_2 by ventilation of the ocean's deepest water. *Paleoceanography* **14**, 571–588 (1999).
- Martin, J. H. Glacial-interglacial CO_2 change: the iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
- Stephens, B. B. & Keeling, R. F. The influence of Antarctic sea ice on glacial-interglacial CO_2 variations. *Nature* **404**, 171–174 (2000).
- Najjar, R. & Orr, J. Design of OCMIP-2 simulations of chlorofluorocarbons, the solubility pump and common biogeochemistry. (<http://www.cgd.ucar.edu/oce/OCMIP/design.pdf>) (1998).
- Gnanadesikan, A., Slater, R. D., Gruber, N. & Sarmiento, J. L. Oceanic vertical exchange and new production: a comparison between models and observations. *Deep Sea Res.* **49**, 363–401 (2002).
- Gnanadesikan, A., Slater, R. D. & Samuels, B. L. Sensitivity of water mass transformation and heat transport to subgrid-scale mixing in coarse-resolution ocean models. *Geophys. Res. Lett.* **30**, 1967, doi:10.1029/2003GL018036 (2003).
- Speer, K., Rintoul, S. R. & Sloyan, B. The diabatic Deacon cell. *J. Phys. Oceanogr.* **30**, 3212–3222 (2000).
- Sloyan, B. M. & Rintoul, S. R. The Southern Ocean limb of the global deep overturning circulation. *J. Phys. Oceanogr.* **31**, 143–173 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I.M. was supported by the DOE Office of Science while at Princeton University, and by the NOAA Postdoctoral Program in Climate and Global Change, administered by the University Corporation for Atmospheric Research, while at MIT. We thank R. Slater for help with the Princeton GCM, D. Sigman and M. Follows for discussions, and R. Anderson for comments that improved the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.M. (imarinov@mit.edu).

LETTERS

Interseismic strain accumulation and the earthquake potential on the southern San Andreas fault system

Yuri Fialko¹

The San Andreas fault in California is a mature continental transform fault that accommodates a significant fraction of motion between the North American and Pacific plates. The two most recent great earthquakes on this fault ruptured its northern and central sections in 1906 and 1857, respectively. The southern section of the fault, however, has not produced a great earthquake in historic times (for at least 250 years). Assuming the average slip rate of a few centimetres per year, typical of the rest of the San Andreas fault, the minimum amount of slip deficit accrued on the southern section is of the order of 7–10 metres, comparable to the maximum co-seismic offset ever documented on the fault^{1,2}. Here I present high-resolution measurements of interseismic deformation across the southern San Andreas fault system using a well-populated catalogue of space-borne synthetic aperture radar data. The data reveal a nearly equal partitioning of deformation between the southern San Andreas and San Jacinto faults, with a pronounced asymmetry in strain accumulation with respect to the geologically mapped fault traces. The observed strain rates confirm that the southern section of the San Andreas fault may be approaching the end of the interseismic phase of the earthquake cycle.

The absence of great historic earthquakes on the southern San Andreas fault (SAF) implies three possibilities. First, the fault may undergo a substantial creep, at least near the surface^{3,4}, similar to the central 40-km-long fault section between Parkfield and San Juan Batista (see Supplementary Fig. 1). Second, the slip rate on the southern SAF may be significantly lower than the rate of 30–40 mm yr⁻¹ measured elsewhere on the SAF⁵, perhaps owing to transfer of slip onto neighbouring faults—in particular, the San Jacinto fault to the west of the SAF⁶. In this case, the seismic slip deficit on the southern SAF may be of the order of a few metres, comparable to the present day slip deficits on the northern and central sections of the SAF. Third, the southern SAF may be accumulating elastic strain at a high (~30 mm yr⁻¹) slip rate, and be late in the interseismic phase of the earthquake cycle. Distinguishing between these possibilities is important, as they have very different implications for the seismic hazard estimates. For example, owing to a lack of significant (moment magnitude greater than 7) historic earthquakes on the southern SAF, and on portions of the San Jacinto fault (for example, the Anza gap), these faults are currently believed to pose the largest seismic risk in California^{2,7}. Except for the first possibility above, the continued quiescence increases the likelihood of a future event.

Different scenarios of strain accumulation and the likelihood of large future events on the southern SAF system may be ultimately discriminated with the help of precise, spatially dense measurements of surface deformation. Previous studies have attempted to estimate the contemporaneous slip rates on the southern SAF system by fitting elastic halfspace models to rather sparse ground-based geodetic measurements from a dozen monuments spanning a ~100-km-long profile across the fault^{8,9}. Also, InSAR (Interferometric Synthetic

Aperture Radar) data collected in the near field of the geologically mapped fault trace were used to detect the presence and extent of fault creep⁴. The far-field point measurements, including GPS (Global Positioning System) and EDM (electronic distance measurements) as well as geologic data, indicate that the southern SAF system accommodates a substantial fraction of the relative motion between the North American and Pacific plates, although the inferred slip rates vary greatly, from as low as 10 mm yr⁻¹ to as high as 35 mm yr⁻¹ (refs 6–10). In this Letter I investigate the interseismic strain accumulation on the southern SAF system using the spatially and temporally dense InSAR data collected by the European Space Agency satellites ERS-1 and ERS-2 between 1992 and 2000, and point measurements of surface velocities derived from the GPS and EDM data representing a time interval between 1985 and 2005.

Figure 1 shows the satellite line-of-sight (LOS) velocity, in millimetres per year, from ERS track 356 covering the San Bernardino-Coachella Valley segment of the southern SAF. The average LOS velocities are derived from a stack of 35 radar interferograms. Details of the InSAR data processing and reduction are presented in Supplementary Information. To account for any residual error due to imprecise orbit information, I subtract the best-fitting linear trend (a planar ‘ramp’) from the interferometric stack using independent data from more than 50 GPS stations located within the radar swath¹¹. The remaining nonlinear signal reveals gradual changes in the radar range that follow the strike of the SAF, and are consistent with the sense of motion between the North American and Pacific plates. Assuming that this signal is due to surface motion, and that the motion is predominantly horizontal and parallel to a local strike of the SAF, the inferred LOS velocity of 13–14 mm yr⁻¹ corresponds to a horizontal velocity of ~45 mm yr⁻¹, in good agreement with the plate tectonic models¹² and GPS measurements, but somewhat higher than the geologic estimates based on cumulative slip rates of major faults in southern California^{6,13}.

It is instructive to compare the InSAR data to independent point measurements of surface deformation from ground-based surveys. Figure 2 shows a comparison between the InSAR data from the A–A’ profile in Fig. 1, and EDM and GPS measurements projected onto the satellite LOS. Locations of the EDM and GPS monuments within the profile are indicated by the respective coloured symbols in Fig. 1. To make the comparison more stringent, the GPS data were not used to remove a residual orbit error from the stacked InSAR data shown in Fig. 2. Instead, the InSAR data were de-trended by fitting a plane to the northeastern part of the interferogram that covers a presumably undeformed part of the North American plate (Fig. 1). The best-fitting plane was then subtracted from the entire interferogram, effectively producing a map of LOS velocities with respect to stable North America. As can be seen from Fig. 2, the InSAR data are in excellent agreement with the GPS and EDM measurements on both ends of the profile, yielding a total relative motion between the North American and Pacific plates in southern California at an average rate

¹Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA.

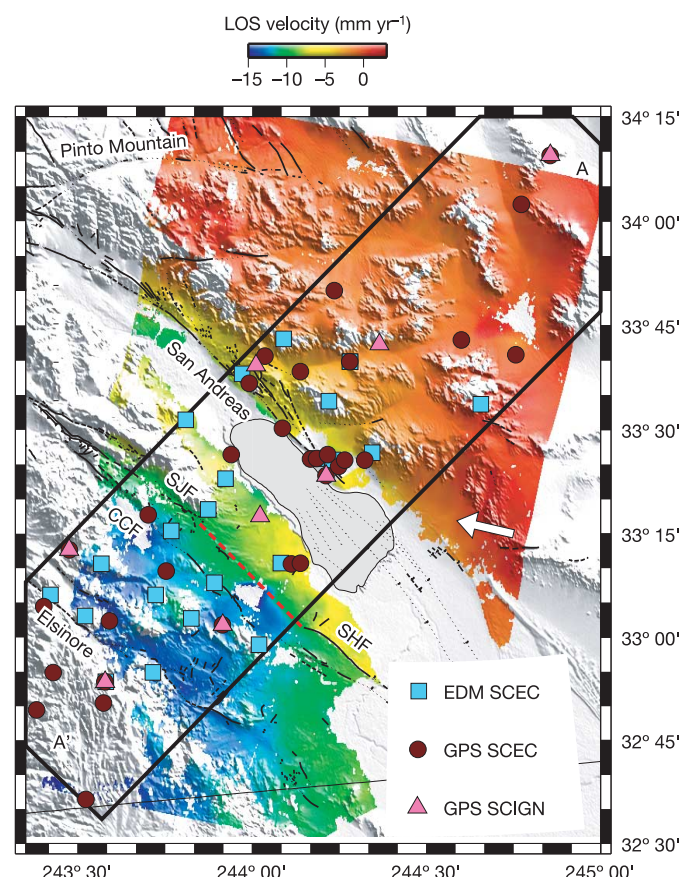


Figure 1 Line of sight (LOS) velocity of the Earth's surface from a stack of radar interferograms spanning a time interval between 1992 and 2000. The velocity map is draped on top of shaded topography. LOS velocities toward the satellite are assumed to be positive. White arrow shows the radar look direction. Black broken lines denote Quaternary faults (SJF, San Jacinto fault; CCF, Coyote Creek fault; SHF, Superstition Hills fault). Red dashed line shows a hypothesized location of an active southern branch of the San Jacinto fault. Black box outlines a profile (A–A') from which the InSAR and GPS data are extracted for a comparison and modelling (Fig. 2). Coloured symbols denote positions of the GPS (Global Positioning System) and EDM (electronic distance measurement) sites within the profile. Data from these sites are provided by the Southern California Earthquake Center (SCEC) and Southern California Integrated GPS Network (SCIGN).

of $45 \pm 2 \text{ mm yr}^{-1}$. All available geodetic data clearly indicate two major zones of high strain rate associated with the San Andreas and San Jacinto faults, and no significant strain accumulation due to the Elsinore fault.

A prominent feature of strain accumulation on the southern San Andreas and San Jacinto faults that was not recognized in the ground data^{8,9} is a significant asymmetry in the velocity gradient with respect to the fault traces. In both cases, velocities on the eastern side of the faults are markedly higher than velocities on the western side. Such an asymmetry is not expected for vertical strike-slip faults provided that the Earth's crust is transversely homogeneous and isotropic. Indeed, no simple dislocation models of interseismic strain accumulation are able to explain the space geodetic data shown in Fig. 2 (see Supplementary Fig. 2). Possible explanations for the concentrated deformation on one side of a fault include across-fault contrasts in the effective shear modulus of the host rocks^{14,15}, postseismic relaxation in the presence of lateral variations in the effective viscosity of the substrate¹⁶, multiple sub-parallel shear zones¹⁷, and a non-vertical fault geometry^{18,19}. Postseismic transients are an unlikely explanation, given the large time lapse since the presumed last great earthquake on the southern SAF.

To test the hypothesis of a rigidity contrast between rocks on different sides of the faults, I performed a number of theoretical simulations of interseismic strain accumulation on the San Andreas–San Jacinto fault system in the presence of lateral variations in the effective shear modulus of the crust. In these simulations I prescribe various slip rates, fault locking depths, and rigidity contrasts between the fault-bounded crustal blocks, and compute the resulting deformation at the Earth's surface. The faults are assumed to be vertical, and follow the geologically mapped traces; for the southern San Jacinto fault, this implies that its currently active branch is the Coyote Creek fault (Fig. 1). Simulations are implemented using the finite element code ABAQUS. The computational domain is a parallelepiped having dimensions of $200 \times 100 \times 30 \text{ km}$ representing the across-fault, depth, and along-fault coordinates, respectively. The interseismic loading is introduced by prescribing a constant strike-slip displacement on a deep extension of the San Andreas and San Jacinto faults. The red line in Fig. 2 shows the theoretical LOS velocities from the best-fitting model.

The best-fitting model suggests a slip rate of 25 mm yr^{-1} and a locking depth of 17 km for the southern SAF, and 21 mm yr^{-1} and 12 km, respectively, for the San Jacinto fault. For both faults, the shear modulus of the western side of the fault is inferred to be three times larger than the shear modulus of the eastern side. There is a certain trade-off between the assumed slip rates, fault locking depths, and rigidity contrasts. On the basis of a number of simulations, the estimated uncertainty in the fault slip rate is $2\text{--}3 \text{ mm yr}^{-1}$, and the uncertainty in the fault locking depth is $2\text{--}4 \text{ km}$. The data require a contrast in the elastic modulus of at least a factor of two, and much higher contrasts (for example, a factor of five) cannot be ruled out. Note that there is some disagreement between the campaign-mode GPS and EDM data from the SCEC (Southern California Earthquake Center) velocity model, on one hand, and the InSAR data and continuous SCIGN (Southern California Integrated GPS Network) GPS data, on the other hand, on the eastern side of the SAF. The SCEC data seem to suggest somewhat lower slip rate and thickness of the brittle layer compared to the InSAR and SCIGN GPS data. This disagreement is unlikely to be due to vertical deformation, as the latter would imply subsidence to the east of the fault. Taking the SCIGN GPS measurements of vertical motion at face value, the data indicate uplift, rather than subsidence, to the east of the fault trace. A good agreement between the InSAR data and well-constrained horizontal velocity from the SCIGN GPS data (see a black triangle at $\sim 105 \text{ km}$ along the profile A–A' in Fig. 2) argues for little, if any, vertical deformation to the east of the fault. Because the best-fitting model is a compromise between all available geodetic data, the deduced slip rate and rigidity contrast on the southern SAF should be considered lower bounds.

Asymmetric patterns of interseismic velocities have been reported elsewhere on the SAF and other major strike-slip faults^{15,20,21}, and often attributed to lateral variations in crustal rigidity^{14,15}. Significant (a factor of two) variations in the effective shear modulus were also reported for kilometre-wide damage zones around several faults in southern California^{11,22}. Variations in the effective elastic properties of the Earth's crust across mature faults are perhaps not surprising, as large-offset faults often juxtapose terrains with dissimilar compositions and mechanical properties²³. However, the inferred magnitude of the rigidity contrasts exceeds high-end estimates from seismic tomography of a factor of $2\text{--}2.5$ (refs 24, 25). The total inferred contrast across the San Andreas and San Jacinto faults is about a factor of $5\text{--}10$, significantly larger than any seismically determined lateral variations in elastic moduli, but comparable to inferences based on geodetic data from other localities¹⁵.

An alternative possibility is that the Quaternary fault traces are not indicative of the fault positions below the brittle layer. To explore this possibility, I hypothesise that the Coyote Creek fault is not the main branch of the southern San Jacinto fault system, and that most of the strain accumulation occurs on an unmapped strike-slip fault

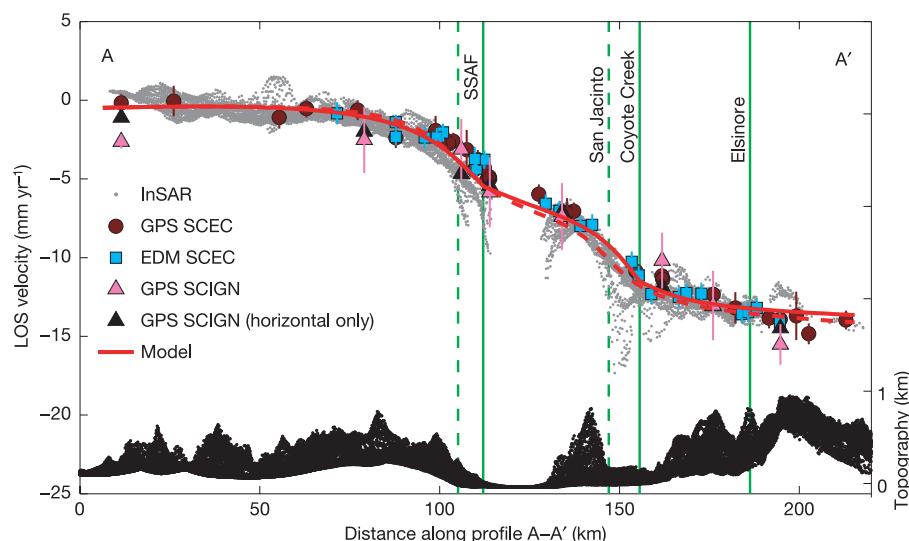


Figure 2 | Average LOS velocities (grey dots) and GPS/EDM data (coloured symbols) projected onto the satellite LOS from the profile A–A' shown in Fig. 1. Vertical bars denote the 2σ errors of the point measurements. Vertical lines denote positions of the mapped fault traces (solid lines), and hypothetical positions of interseismic creep at the bottom of the brittle layer (dashed lines). Solid red line is a theoretical model of interseismic strain

accumulation due to a deep slip below the mapped traces of the southern San Andreas and San Jacinto faults in the presence of lateral variations in the crustal rigidity. Dashed red line is a theoretical model of the same parameter for the inferred alternative fault positions, assuming no lateral variations in the rock rigidity.

connecting the southern termination of the San Jacinto fault and the Superstition Hills fault (see dashed red line in Fig. 1). In the case of the SAF, one might argue that the fault is dipping to the east^{26,27}, based on the fact that seismicity occurs several kilometres off the fault trace²⁸. The hypothesized alternative locations of shear zones driving the interseismic deformation at the brittle–ductile transition are denoted by dashed green lines in Fig. 2. Under these assumptions, it is possible to explain the data equally well without appealing to variations in the effective shear modulus of the crust (see dashed red line in Fig. 2). In this case, the inferred slip rates are 25 mm yr^{-1} and 19 mm yr^{-1} and the locking depths are 12 and 10 km for the San Andreas and San Jacinto faults, respectively.

The assumed position of the southern San Jacinto fault (Fig. 1) is supported by a lineament of microseismicity²⁸, yet the absence of an active fault trace is puzzling. Partly, such absence might be explained by alluvial burial from the ancient Lake Cahuilla²⁹. However, it is not clear whether the corresponding fault segment remained quiescent over the 400 years since the lake retreat²⁹, or if it could be a 'blind' strike-slip fault¹⁹. For the SAF, the fault dip angle required by the homogeneous model is $\sim 60^\circ$. A steeper fault implies a greater rigidity contrast. Ultimately, the proposed interpretations of interseismic strain accumulation admit a robust observational test. The rigidity contrast model predicts that if future major earthquakes occur on subvertical ruptures coincident with the mapped traces of the San Andreas and San Jacinto faults, the resulting coseismic displacement field should be essentially asymmetric, with displacements on the eastern side of a fault being at least two to three times larger than displacements on the western side of a fault. Ruptures having the proposed alternative fault geometries will be direct evidence against large rigidity contrasts. These results suggest that, in general, information about coseismic deformation on a given fault—such as asymmetries in the radiation pattern, and static displacement fields^{18,27}—may greatly reduce uncertainties in the interpretation of the interseismic deformation data.

Regardless of the details of fault geometry and mechanical properties of the ambient crust, results presented in this study lend support to intermediate-term forecasts of a high probability of major earthquakes on the southern SAF system^{2,7}. Space geodetic data shown in Figs 1 and 2 clearly demonstrate that the southern SAF is accumulating significant elastic strain, as indicated by a broad area of high

gradients in the LOS velocity field on the eastern side of the fault. Although some creep may be occurring in the uppermost crust⁴, as evidenced by a steep gradient in the LOS velocity immediately to the west of the surface trace of the SAF (Fig. 2), it does not prevent (and if anything, enhances) a build-up of tectonic stress on the rest of the fault at seismogenic depths. I point out that the step-like increase in the radar range across the SAF may not be entirely due to right-lateral fault creep, and probably involves ground subsidence to the west of the fault. Without subsidence, the observed variations in the LOS velocity would imply left-lateral deformation within the decorrelated area and immediately to the west of Salton Sea (at ~ 120 – 130 km along the profile A–A', see Fig. 2), which is highly unlikely. The inferred ground subsidence may be either man-made (for example, due to agricultural activities in the Coachella Valley), or of tectonic origin (for example, indicating secular deepening of the Salton Trough). Although it might be possible to discriminate between the horizontal and vertical components of deformation using complementary InSAR data from ascending orbits, unfortunately no suitable acquisitions are available from the ERS satellites.

The current slip rate on the southern SAF of $25 \pm 3 \text{ mm yr}^{-1}$ determined from the space geodetic data is in excellent agreement with some long-term geologic estimates (for example, $25 \pm 4 \text{ mm yr}^{-1}$; ref. 13), although other geologic studies may suggest lower rates¹⁰. The discrepancy between different geologic data sets might be due to along-fault variations in the mechanical behaviour of the uppermost crustal layer. For example, zones of apparently low slip rates might represent substantial inelastic yielding of a shallow layer in the interseismic period, either in the form of creep⁴, or more distributed failure¹⁹. It is reasonable to assume that the geodetically determined slip rate remained relatively constant in the recent geologic past, and, in particular, over the past several hundred years. It follows that the relative seismic quiescence of the southern SAF over the past 250 years implies a slip deficit of 5.5–7 m. This may be compared to palaeoseismological estimates of average recurrence times of large events on the southern SAF of 200–300 years, and average coseismic displacements of 4–7 m (refs 1, 2). Although simple time- and size-predictable earthquake models have been shown to be inadequate^{30–32}, and the repeat interval between large earthquakes may vary significantly, it may be argued that the accumulated slip deficit cannot greatly exceed the maximum

interseismic displacement documented throughout the fault history (see figure 10 in ref. 31 for example). If so, the southern SAF is probably in the late phase of its interseismic recurrence^{2,7,9}.

The data and modelling results presented in this study also reveal a surprisingly robust strain accumulation on the San Jacinto fault. The inferred slip rate on this fault of 19–21 mm yr⁻¹ is significantly higher than most current estimates, but is in good agreement with geologic data representing average slip rates over 10⁵–10⁶ yr timescales^{6,33}. The lower (10–12 mm yr⁻¹) geodetic slip rates on the San Jacinto fault inferred from previous GPS studies may be in part due to the use of homogeneous elastic halfspace models (as well as assumed fault locations; see Fig. 2, and Supplementary Fig. 2) for data interpretation. The spatially dense InSAR data demonstrate that the total slip rate on the SAF–San Jacinto system of 45 mm yr⁻¹ is nearly equally partitioned between these two faults (Figs 1 and 2). Together, the San Jacinto fault and the southern SAF appear to accommodate the bulk of the relative motion between the North American and Pacific plates in southern California. These results imply little, if any, deformation on other major crustal faults to the west of San Jacinto.

It should be noted that although the highly accurate and detailed geodetic data provide useful constraints on the rate of the interseismic build-up of stress, and the probability of large events on a given fault, the relationships between loading rate, absolute stress level, and rate of seismicity are still poorly understood. For example, the low-slip-rate Elsinore fault generates significant microseismicity. Intense microseismicity is also associated with the faster-moving San Jacinto fault, while the SAF, characterized by the highest slip rate, is practically devoid of microearthquakes^{25,28}. These patterns exemplify extremely complex relationships between tectonic loading rate and seismic activity.

Received 1 September 2005; accepted 6 April 2006.

- Sieh, K., Stuiver, M. & Brillinger, D. A more precise chronology of earthquakes produced by the San Andreas fault in Southern California. *J. Geophys. Res.* **94**, 603–623 (1989).
- Weldon, R. J., Fumal, T. E., Biasi, G. P. & Scharer, K. M. Geophysics—Past and future earthquakes on the San Andreas fault. *Science* **308**, 966–967 (2005).
- Turcotte, D. L. & Schubert, G. *Geodynamics* 2nd edn (Cambridge Univ. Press, New York, 2002).
- Lyons, S. & Sandwell, D. Fault creep along the southern San Andreas from interferometric synthetic aperture radar, permanent scatterers, and stacking. *J. Geophys. Res.* **108**, doi:10.1029/2002JB001831 (2003).
- Thatcher, W. & Lisowski, M. Long-term seismic potential of the San Andreas fault southeast of San Francisco, California. *J. Geophys. Res.* **92**, 4771–4784 (1987).
- Bennett, R. A., Friedrich, A. M. & Furlong, K. P. Codependent histories of the San Andreas and San Jacinto fault zones from inversion of fault displacement rates. *Geology* **32**, 961–964 (2004).
- Working Group on California Earthquake Probabilities. Seismic hazards in southern California: Probable earthquakes, 1994–2024. *Bull. Seismol. Soc. Am.* **85**, 379–439 (1995).
- Johnson, H. O., Agnew, D. C. & Wyatt, F. K. Present-day crustal deformation in southern California. *J. Geophys. Res.* **99**, 23951–23974 (1994).
- Bennett, R. A., Rodi, W. & Reilinger, R. E. Global positioning system constraints on fault slip rates in southern California and northern Baja, Mexico. *J. Geophys. Res.* **101**, 21943–21960 (1996).
- van der Woerd, J. *et al.* Long-term slip rate of the southern San Andreas Fault, from ¹⁰Be–²⁶Al surface exposure dating of an offset alluvial fan. *J. Geophys. Res.* **111**, B04407, doi:10.1029/2004JB003559 (2006).
- Fialko, Y. Probing the mechanical properties of seismically active crust with space geodesy: Study of the co-seismic deformation due to the 1992 M_w7.3 Landers (southern California) earthquake. *J. Geophys. Res.* **109**, B03307, doi:10.1029/2003JB002756 (2004).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Current plate motions. *Geophys. J. Int.* **101**, 425–478 (1990).
- Weldon, R. J. & Sieh, K. E. Holocene rate of slip and tentative recurrence interval for large earthquakes on the San Andreas fault, Cajon Pass, southern California. *Geol. Soc. Am. Bull.* **96**, 793–812 (1985).
- Lisowski, M., Savage, J. & Prescott, W. H. The velocity field along the San Andreas fault in central and southern California. *J. Geophys. Res.* **96**, 8369–8389 (1991).
- Le Pichon, X., Kreemer, C. & Chamot-Rooke, N. Asymmetry in elastic properties and the evolution of large continental strike-slip faults. *J. Geophys. Res.* **110**, B03405, doi:10.1029/2004JB003343 (2005).
- Li, V. C. & Rice, J. Crustal deformation in great California earthquake cycles. *J. Geophys. Res.* **92**, 11533–11551 (1987).
- Kenner, S. & Segall, P. Lower crustal structure in Northern California: Implications from strain rate variations following the 1906 San Francisco earthquake. *J. Geophys. Res.* **108**, 2011, doi:10.1029/2001JB000189 (2003).
- Fialko, Y., Simons, M. & Agnew, D. The complete (3-D) surface displacement field in the epicentral area of the 1999 M_w7.1 Hector Mine earthquake, southern California, from space geodetic observations. *Geophys. Res. Lett.* **28**, 3063–3066 (2001).
- Fialko, Y., Sandwell, D., Simons, M. & Rosen, P. Three-dimensional deformation caused by the Bam, Iran, earthquake and the origin of shallow slip deficit. *Nature* **435**, 295–299 (2005).
- Prescott, W. H. & Yu, S. B. Geodetic measurement of horizontal deformation in the Northern San Francisco Bay region, California. *J. Geophys. Res.* **91**, 7475–7484 (1986).
- Freymueller, T. T., Murray, M. H., Segall, P. & Castillo, D. Kinematics of the Pacific North America plate boundary zone, northern California. *J. Geophys. Res.* **104**, 7419–7441 (1999).
- Fialko, Y. *et al.* Deformation on nearby faults induced by the 1999 Hector Mine earthquake. *Science* **297**, 1858–1862 (2002).
- Thatcher, W. Microplate versus continuum description of active tectonic deformation. *J. Geophys. Res.* **100**, 3885–3894 (1983).
- Ben-Zion, Y. & Andrews, D. J. Properties and implications of dynamic rupture along a material interface. *Bull. Seismol. Soc. Am.* **88**, 1085–1094 (1998).
- Hauksson, E. Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California. *J. Geophys. Res.* **105**, 13875–13903 (2000).
- Yule, D. & Sieh, K. Complexities of the San Andreas fault near San Geronio Pass: Implications for large earthquakes. *J. Geophys. Res.* **108**, doi:10.1029/2001JB000451 (2003).
- Fialko, Y., Rivera, L. & Kanamori, H. Estimate of differential stress in the upper crust from variations in topography and strike along the San Andreas fault. *Geophys. J. Int.* **160**, 527–532 (2005).
- Richards-Dinger, K. & Shearer, P. Earthquake locations in southern California obtained using source specific station terms. *J. Geophys. Res.* **105**, 10939–10960 (2000).
- Waters, M. R. Late Holocene lacustrine chronology and archaeology of ancient lake Cahuilla, California. *Quat. Res.* **19**, 373–387 (1983).
- Murray, J. & Segall, P. Testing time-predictable earthquake recurrence by direct measurement of strain accumulation and release. *Nature* **419**, 287–291 (2002).
- Weldon, R. J., Scharer, K. M., Fumal, T. E. & Biasi, G. P. Wrightwood and the earthquake cycle: What a long recurrence record tells us about how faults work. *GSA Today* **14**, 4–10 (2004).
- Cisternas, M. *et al.* Predecessors of the giant 1960 Chile earthquake. *Nature* **437**, 404–407 (2005).
- Sharp, R. V. San Jacinto fault zone in the Peninsular Ranges of southern California. *Geol. Soc. Am. Bull.* **78**, 705–730 (1967).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank R. Weldon and P. Segall for suggestions. This work was supported by the NSF and the Southern California Earthquake Center (SCEC). Original InSAR data are copyright of the European Space Agency, distributed by Eurimage, Italy, and acquired via the WInSAR Consortium. The ERS SAR imagery was processed using the JPL/Caltech software package ROI_PAC. The continuous GPS data were provided by the Scripps Orbit and Permanent Array Center (SOPAC), and the campaign GPS and EDM data were provided by the Crustal Motion Model v3 of SCEC.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (yfialko@ucsd.edu).

LETTERS

A lamprey from the Cretaceous Jehol biota of China

Mee-mann Chang¹, Jiangyong Zhang¹ & Desui Miao²

Widespread nowadays in freshwater and coastal seas of the cold and temporal zones, lampreys are a jawless vertebrate group that has been in existence for more than 300 million years but left a meagre fossil record. Only two fossil lamprey species, namely *Mayomyzon pieckoensis*^{1,2} and *Hardistiella montanensis*^{3–5}, have been recognized with certainty from North American Carboniferous marine deposits⁶. Here we report a freshwater lamprey from the Early Cretaceous epoch (about 125 million years ago) of Inner Mongolia, China. The new taxon, *Mesomyzon mengae*, has a long snout, a well-developed sucking oral disk, a relatively long branchial apparatus showing branchial basket, seven gill pouches, gill arches and impressions of gill filaments, about 80 myomeres and several other characters that are previously unknown or ambiguous. Our finding not only indicates *Mesomyzon*'s closer relationship to extant lampreys but also reveals the group's invasion into a freshwater environment no later than the Early Cretaceous. The new material furthers our understanding of ancient lampreys, bridges the gap between the Carboniferous ones and their recent relatives, and adds to our knowledge of the evolutionary history of lampreys.

Class Agnatha

Order Petromyzontiformes (Hyperoartii)

Family Petromyzontidae

Mesomyzon mengae gen. et sp. nov.

Holotype. IVPP (Institute of Vertebrate Paleontology and Paleoanthropology, Beijing, China) collection number V14719, a nearly complete adult specimen, without preservation of the tail.

Referred material. IVPP V14718A, B (part and counterpart), a complete specimen with the anterior tip of the snout unpreserved.

Etymology. The generic name derives from meso (Greek): Mesozoic; myzon (Greek), sucker. The species name honours Meng Qingwen from the University of Fisheries, Shanghai, whose lifelong studies on Chinese fishes have inspired us.

Horizon and locality. Yixian Formation, Lower Cretaceous; Ningcheng, Inner Mongolia, China.

Diagnosis. A small lamprey, body elongated, eel-shaped, length 12 times its depth, 4 times its head length; preorbital region fairly long, nearly one-third of the head length, 5 times the eye diameter; oral opening surrounded by radiating rectangular depressed areas, presumably covered by tooth plates with teeth, forming sucking disk; branchial basket well developed, with seven gill pouches, branchial apparatus obviously longer than preorbital region; first gill pouch posteroventral to otic capsule; eight or nine preserved gonads of rounded shape, non-metameric; trunk muscle segments more than 80; paired fins and anal fin absent, dorsal fin above posterior portion of body, caudal fin diphyccercal.

The two specimens of the new lamprey *Mesomyzon* are nearly of the same size. The body is elongated and eel-shaped (Fig. 1a–c). The anterior portion of the body is rounded, whereas the tail is more compressed laterally and tapers posteriorly. The body length

measured in the holotype of *Mesomyzon* V14719 (with the tail missing) is about 83.5 mm, whereas in V14718B (without the tip of the snout) it is 85 mm. Thus, the specimens of *Mesomyzon* are longer than those of *Mayomyzon* (33–61 mm) from Illinois^{1,2}, but slightly shorter than the complete one of the three known *Hardistiella* specimens ("which does not exceed 10 cm in total length"⁵) from Montana, and much shorter than most recent lampreys^{7,8}. Its ratio of body length to body depth is about 12 (on the basis of both V14718B and V14719). The corresponding ratio is 8 to 9 in *Mayomyzon*¹, 10 in *Hardistiella* (inferred from Fig. 1A of ref. 2) and 12–20 in the recent lampreys from China⁸ and elsewhere in the world⁷. *Mesomyzon* is thus obviously slenderer than the two Carboniferous lampreys and closer in body proportion to the living ones. In the holotype in which the head is well preserved, the ratio of body length to head length (measured from the tip of the snout to the posterior margin of the last gill pouch) is about 4, approaching that of the recent *Lampetra* from China (4.5–6.0)⁸. The snout (or preorbital region) is well developed. The head length is roughly 2.8 times the preorbital length in the holotype, falling in the range of the recent *Lampetra* from China (2.7–3.4)⁸. The length of the snout is five times of the eye diameter in the same specimen, and much longer in proportion than that in *Mayomyzon*, namely 3.5 (ref. 1).

The holotype lies, for the most part, on its flank but with its head and anterior portion of the body rotated (along its axis) slightly towards its left so that the right side of the head is turned somewhat upwards. As a result of the rotation, the right eye, as a round dark stain (i.e., Fig. 1d, e), is situated near the top line of the head, whereas the left eye, as a faint circle with rough interior (r.e., Fig. 1d, e), is situated under, and partly overlapped by, the right eye. A small circle in the middle of the bigger circle of the right eye may indicate the presence of the lens. The left and right otic capsules (l.ot. and r.ot., Fig. 1d, e) are just behind the eyes and displayed in the same way as the eyes. The nasohypophysial opening was not seen in either of the specimens because of the state of preservation.

No parts or imprints of the endoskeleton of the cranium could be observed because it was neither calcified nor ossified. However, it is extremely fortunate that, on the holotype, a subterminal sucking oral disk (or.d., Fig. 1d, e) is clearly shown in the form of depressed rectangular areas divided by ridges, arranged in radiating rows around the oral opening. This is somewhat similar to the mouth structure of the enigmatic lamprey-like *Pipiscius* (see Figs 3–5 in ref. 9) from the Carboniferous period of Illinois^{10,11}. We could count four or five such areas on the nearly half of the oral disk preserved on the holotype. These depressed areas presumably carried tooth plates with teeth when the animal was alive¹², although no horny teeth were found on the specimens. If this were so, the oral tooth plates in *Mesomyzon* must be much bigger, and fewer in number, than in *Pipiscius*⁹ and recent lampreys^{7,8,12}.

The usually cartilaginous visceral skeleton in the form of the branchial basket (br.b., Fig. 1d, e) shows its distinct general outline and structure in light-grey stain on the holotype. The branchial

¹Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, PO Box 643, Beijing 100044, China. ²Natural History Museum and Biodiversity Research Center, University of Kansas, Lawrence, Kansas 66045, USA.

basket of lamprey has not previously been observed in fossils, although elongated branchial baskets with many branchial units have been seen in the Late Devonian anaspid-like form *Euphanerops longaevus* from Miguasha, Québec, Canada^{5,10}. Not only are the outlines of the seven gill pouches (g.p., Fig. 1d, e) clearly visible but the impressions of the gill filaments (g.f., Fig. 1d, e) can also be traced in many places posterior to the vertical bars (gill arches, g.a., Fig. 1d, e) of the branchial basket. The length of the branchial apparatus obviously exceeds that of the preorbital region, whereas the gill pouches in *Mayomyzon* and *Hardistiella* are closely set and the branchial apparatus is hardly longer than the preorbital region^{4,10}. The first gill pouch is situated posteroventral to the otic capsule, as in recent lampreys but in sharp contrast to *Mayomyzon*^{1,2} and *Hardistiella*³, in which it is situated close behind the eyes. Unfortunately, no distinct gill openings could be detected on the specimens, although we believe there should be seven separate ones because in other respects *Mesomyzon* is too similar to the extant lamprey. Posterior to the last gill pouch, faintly visible are the upper and lower parts of the anterior portion of the pericardial cartilage (p.c., Fig. 1d, e). Two thin lines, stretching forwards from the anterior part of the branchial apparatus, seem to enclose a rod-like structure. Conceivably, it could be interpreted as the piston cartilage (p.c.?, Fig. 1d, e) that supports the rasping tongue. Behind the anterior part

of the pericardial cartilage, the outline of the liver (l., Fig. 1d, e) can also be traced. As judged from the situation in recent lampreys, the liver might be longer than that in our illustration. The digestive tract (d.t.?, Fig. 1d, e) is probably represented by a thin dark band that extends backwards from the end of the branchial basket, along the upper border of the liver, to an area below the dorsal fin, where it bends down to the probable anus (a., Fig. 1c). However, the interpretation of the thin dark band as the dorsal aorta is not ruled out.

Filled with coarse sediments, lined in a row and situated in the posterior half of the abdomen are eight or nine circular spaces, here interpreted as gonads (g., Fig. 1c), although they might alternatively be interpreted as gut. They are not arranged metamerically. Lamprey females have a single horseshoe-shaped ovary, whereas the testis of males consists of leaf-like lobes^{7,13,14}. The appearance of the gonads in the new form seems similar to that in the latter. Numerous myosepta (ms., Fig. 1c) are distinctly visible in the holotype. About 80 or more trunk myomeres can be counted. Under the dorsal margin of the body and parallel to the margin is the notochord (nc., Fig. 1c), a dark-stained thick band that stretches from the head region back to the end of the body. The traceability of the notochord in the fossils of many soft-bodied animals is certainly due to the more decay-resistant characteristic of its sheath¹⁵. The myoseptum that crosses

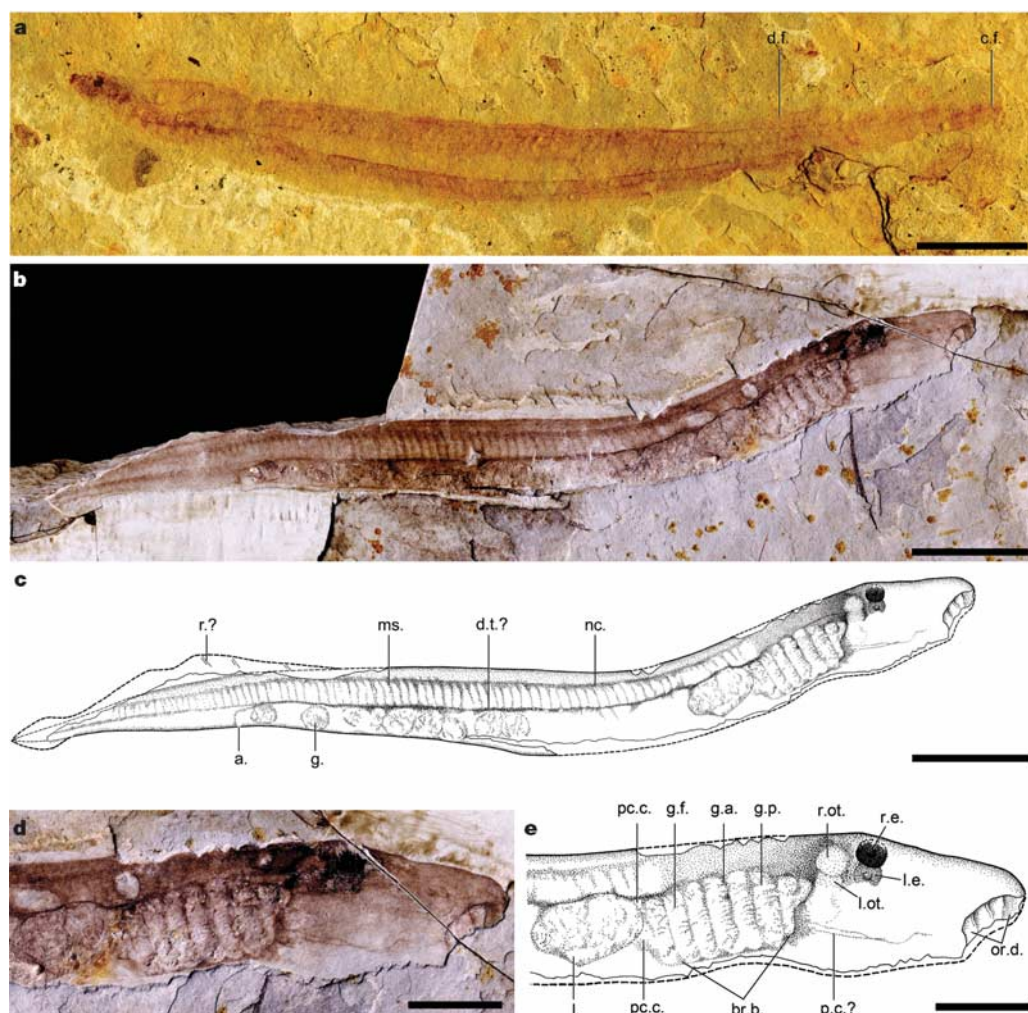


Figure 1 | *Mesomyzon mengae* gen. et sp. nov. **a**, A complete fish (IVPP V14718A) in left view. **b**, Holotype (IVPP V14719) in right view. **c**, Drawing of the holotype, with the dorsal fin and caudal region reconstructed on the basis of IVPP V14718A. **d**, Photograph of head and anterior part of body of the holotype. **e**, Drawing of the same part as in **d**. Scale bars, 10 mm (**a–c**)

and 5 mm (**d, e**). Abbreviations: a., anus; br.b., branchial basket; c.f., caudal fin; d.f., dorsal fin; d.t.?, possible digestive tract; g., gonads; g.a., gill arches; g.f., gill filaments; l., liver; l.e., left eye; l.ot., left otic capsule; ms., myosepta; nc., notochord; or.d., oral disk; p.c.?, possible piston cartilage; p.c., pericardial cartilage; r.e., right eye; r.ot., right otic capsule.

the notochord in each segment made it difficult to observe the shape of arcualia, if any is preserved there. Paired fins and anal fin are absent. The dorsal fin (d.f., Fig. 1a) is seen in V14718B above the posterior portion of the body. It most probably merges with the caudal fin (c.f., Fig. 1a). Very faint impressions of rays are occasionally seen in the dorsal fin in V14718B (r.f., Fig. 1c). The tail seems diphyrcal rather than hypocercal on the same specimen.

Lacking mineralized tissues such as bone or calcified cartilage, lampreys are seldom preserved as fossils; their early evolutionary history is therefore still poorly known. *Mesomyzon* is the third fossil genus of the group ever discovered and in some respects is the best preserved. It lived more than 100 Myr ago, about 200 Myr more recent than their known precursors from the Carboniferous. The lamprey larvae—ammocoetes—are different from the adults. They are blind, with a short snout bending down, and without a sucking disk^{7,8,10}. Equipped with well-defined eyes, a long snout and a well-developed sucking oral disk, *Mesomyzon* is undoubtedly either an adult or a late transformer, instead of a larval form.

Mesomyzon possesses a slenderer body than its Carboniferous relatives, with the length-to-depth proportion close to that of the recent lampreys. Its snout is long and its oral sucker well developed. The structure of its branchial basket and pericardial cartilage is almost the same as in the recent lampreys. There are seven gill pouches. The ratios of the length of the branchial apparatus to that of the preorbital region are also intermediate between the Carboniferous precursors and the extant lampreys. Its first gill pouch is situated posteroventral to the otic capsule instead of close behind the eyes as in *Mayomyzon*^{1,2} and *Hardistiella*³. The distinctly shown trunk muscle segments are numerous (more than 80). Many of these characters have not previously been demonstrated in fossils. They suggest that *Mesomyzon* is unquestionably more closely related to the recent lampreys than to its Carboniferous kin. Consequently, the lamprey has already approximated to its current appearance by the Early Cretaceous. It therefore provides another interesting example from the fossil record in the discussion of rates of evolution. For about 100 Myr, lampreys underwent few morphological changes and therefore showed unusual morphological conservatism, or evolutionary stasis.

In living sea lampreys the teeth are numerous and are arranged in radiating rows, whereas in most river and brook lampreys they are not in radiating rows but in a more complicated pattern^{7,8,12}. The oral disk of *Mesomyzon* with radiating depressed areas (apparently covered by tooth plates) is more similar to the pattern of a sea lamprey. The development of an oral sucking disk brought about the possibility of a specific suctorial, parasitic mode of life. The material of *Mesomyzon* was collected from the freshwater shale deposits of the Yixian Formation. The new locality is not far from the famous Jehol Biota¹⁶ localities in Liaoning with its present latitude of about 41° N, which is within the range of the palaeolatitude of the Jehol Biota (40–45° N (ref. 17)) during the Early Cretaceous. Associated with the new lamprey are abundant fossils commonly found in the Jehol Biota, including insects, a teleost *Lycoptera davidi* (Osteoglossomorpha), a salamander, a lizard and an unidentified fossil bird. All of these animals are terrestrial or freshwater dwellers. There is also no indication whatsoever that the area has had any connection with

the sea since the Triassic period¹⁸. *Mesomyzon* was no doubt also a freshwater dweller. It probably evolved from an anadromous fore-runner because it still kept the basic pattern of an oral sucker characteristic of a sea lamprey. It seems likely that living lampreys originated from anadromous ancestors before some of them became landlocked in the freshwater environment.

Received 13 January; accepted 13 March 2006.

1. Bardack, D. & Zangerl, R. First fossil lamprey: A record from the Pennsylvanian of Illinois. *Science* **162**, 1265–1267 (1968).
2. Bardack, D. & Zangerl, R. in *The Biology of lampreys* Vol. 1 (eds Hardisty, M. W. & Potter, I. C.) 67–84 (Academic, London, 1971).
3. Janvier, P. & Lund, R. *Hardistiella montanensis* n. gen. et sp. (Petromyzontida) from the Lower Carboniferous of Montana, with remarks on the affinities of the lampreys. *J. Vertebr. Paleontol.* **2**, 407–413 (1983).
4. Lund, R. & Janvier, P. A second lamprey from the Lower Carboniferous (Namurian) of Bear Gulch, Montana (U.S.A.). *Geobios* **19**, 647–652 (1986).
5. Janvier, P., Lund, R. & Grogan, E. Further consideration of the earliest known lamprey *Hardistiella montanensis* Janvier and Lund, 1983, from the Carboniferous of the Bear Gulch, Montana, U.S.A. *J. Vertebr. Paleontol.* **24**, 742–743 (2004).
6. Grogan, E. D. & Lund, R. The geological and biological environment of the Bear Gulch limestone (Mississippian of Montana, USA) and a model for its deposition. *Geodiversitas* **24**, 295–315 (2002).
7. Hardisty, M. W. & Potter, I. C. (eds). *The Biology of Lampreys* (Academic, London, 1974–82).
8. Zhu, Y. D. et al. *Fauna Sinica: Cyclostomata, Chondrichthyes* (Science Press, Beijing, 2001).
9. Bardack, D. & Richardson, E. S. Jr. New agnathous fishes from the Pennsylvanian of Illinois. *Fieldiana Geol.* **33**, 489–510 (1977).
10. Janvier, P. *Early Vertebrates* (Clarendon, Oxford, 1996).
11. Janvier, P. The dawn of the vertebrates: Character versus common ascent in the rise of current vertebrate phylogenies. *Palaeontology* **39**, 259–287 (1996).
12. Potter, I. C. & Hilliard, R. W. A proposal for the functional and phylogenetic significance of differences in the dentition of lampreys (Agnatha: Petromyzontiformes). *J. Zool.* **212**, 713–737 (1987).
13. Barker, L. A., Morrison, B. J., Wicks, B. J. & Beamish, F. W. H. Potential fecundity of landlocked sea lamprey larvae, *Petromyzon marinus*, with typical and atypical gonads. *Copeia* **1998**, 1070–1075 (1998).
14. Wicks, B. J., Barker, L. A., Morrison, B. J. & Beamish, F. W. H. Gonadal variation in Great Lakes sea lamprey, *Petromyzon marinus*, larvae. *J. Great Lakes Res.* **24**, 962–968 (1998).
15. Briggs, D. E. G. & Kear, A. J. Decay of *Branchiostoma*: Implications for soft-tissue preservation in conodonts and other primitive chordates. *Lethaia* **26**, 275–287 (1994).
16. Chang, M. M., Chen, P. J., Wang, Y. Q., Wang, Y. & Miao, D. (eds). *The Jehol Biota* (Shanghai Scientific and Technical Publishers, Shanghai, 2003).
17. Zhou, Z., Barrett, P. M. & Hilton, J. An exceptionally preserved Lower Cretaceous ecosystem. *Nature* **421**, 807–814 (2003).
18. Wang, H. Z. (ed.) *Atlas of the Palaeogeography of China* (Cartographic Publishing House, Beijing, 1985).

Acknowledgements We thank X. Wang for providing the fossils and information on their provenance and stratigraphic horizon; Z. Zhou for discussions and reading the manuscript; M. Zhu for providing information and discussions; M. Yang for the illustrations; and W. Gao for photography. This study was supported by the Chinese Academy of Sciences, the National Natural Science Foundation of China and the Asian–Swedish Research Partnership Program of the Swedish Research Council.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.-M.C. (zhangmiman@ivpp.ac.cn or cmeemann@yahoo.com).

Image scoring and cooperation in a cleaner fish mutualism

Redouan Bshary¹ & Alexandra S. Grutter²

Humans are highly social animals and often help unrelated individuals that may never reciprocate the altruist's favour^{1–5}. This apparent evolutionary puzzle may be explained by the altruist's gain in social image: image-scoring bystanders, also known as eavesdroppers, notice the altruistic act and therefore are more likely to help the altruist in the future^{5–7}. Such complex indirect reciprocity based on altruistic acts may evolve only after simple indirect reciprocity has been established, which requires two steps. First, image scoring evolves when bystanders gain personal benefits from information gathered, for example, by finding cooperative partners^{8–10}. Second, altruistic behaviour in the presence of such bystanders may evolve if altruists benefit from access to the bystanders. Here, we provide experimental evidence for both of the requirements in a cleaning mutualism involving the cleaner fish *Labroides dimidiatus*. These cleaners may cooperate and remove ectoparasites from clients or they may cheat by feeding on client mucus^{11,12}. As mucus may be preferred over typical client ectoparasites¹³, clients must make cleaners feed against their preference to obtain a cooperative service. We found that eavesdropping clients spent more time next to 'cooperative' than 'unknown cooperative level' cleaners, which shows that clients engage in image-scoring behaviour. Furthermore, trained cleaners learned to feed more cooperatively when in an 'image-scoring' than in a 'non-image-scoring' situation.

Early game theoretic models of indirect reciprocity based on image scoring assumed that every single act is altruistic and that image scoring is already present in a population, as is the strategy to help individuals that have been observed to help others^{5–7}. Thus, these models do not address why image scoring and helping should evolve in the first place.

Communication network theory¹⁴, however, provides a framework for indirect reciprocity based on image scoring. Image scoring, called eavesdropping in the communication literature¹⁵, involves collecting information on interactions between third parties where the information gained provides direct benefits to the bystander. This image-scoring behaviour in bystanders leads to changes in their behaviour, best documented in situations involving conflict, where individuals may adjust their fighting behaviour according to knowledge gained by observing their opponent^{16,17}. Thus, bystanders attribute an image score about fighting parameters to observed individuals. Eavesdropping in agonistic contexts has been documented across a wide range of animal taxa¹⁵. In the context of cooperation, image scoring may evolve if individuals differ consistently in the amount that they cooperate^{10,18}. Here, observers may select individuals as partners for cooperative tasks based on how cooperative they are in their current interaction. Once such egoistic image scoring/eavesdropping has evolved, interacting individuals are selected to adjust appropriately their behaviour according to the presence or absence of bystanders (termed "audience effects"¹⁹).

Therefore, in the context of cooperation, image-scoring observers would select for cooperative behaviour as an audience effect, as it would provide the altruist with access to a cooperating partner.

Altruism based on image scoring and indirect reciprocity may occur in symmetric games (prisoner's dilemma-type games) where both partners may cooperate or cheat. However, the game may also be asymmetric where only one class of players can choose between helping and cheating its partner, while the other class of players collects crucial information by image scoring. For example, males may use helping as a behavioural handicap while females are the eavesdroppers^{8–10}. In addition, many interspecific mutualisms consist of asymmetric investment by one player while the returns of the partner come as a by-product²⁰—cleaning mutualism is a typical example²¹.

The interactions between the cleaner fish *Labroides dimidiatus* and its client fishes are a well-known example of interspecific mutualism²². Cleaners may cooperate and remove ectoparasites from client fish, or they may cheat by feeding on client mucus^{11,12}, which they prefer over typical client ectoparasites¹³. Hence, there is a temptation for cleaners to cheat. Although a few piscivorous clients could cheat by attempting to eat the cleaner fish²², the vast majority of clients consist of species that do not predate on fish²³. Such clients have no means to exploit a cleaner fish, precluding tit-for-tat-like²⁴ control strategies. Instead, field observations suggest that image scoring is one of several alternative mechanisms²¹ used by clients to avoid exploitation by cheating cleaners. Client fish almost always invite a

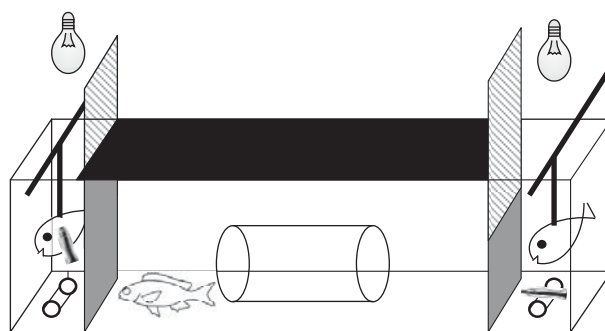


Figure 1 | Experimental set-up to test for image scoring. Aquarium (95 × 36 × 36 cm) with a black cloth over the middle and 40-W incandescent lights over the side compartments, enhancing the one-way mirrors, which allowed the client in the middle to see the cleaner fish and client model in the side compartments. Pipes served as shelters for fish. Cooperative cleaners foraged on prawn (0.002 g) smeared on the model, whereas unknown cooperation level cleaners swam around freely. The time the client spent next to each cleaner with its entire body out of the shelter was recorded. Hatched area indicates opaque sliding covers.

¹University of Neuchâtel, Department of Zoology, Rue Emile-Argand 11, Case postale 158, 2009 Neuchâtel, Switzerland. ²University of Queensland, School of Integrative Biology, St Lucia QLD 4072, Australia.

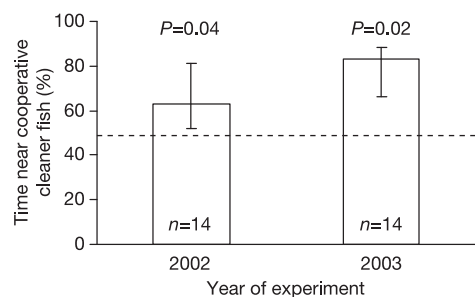


Figure 2 | Client fish conduct image scoring. The percentage of time a client fish spent near the cooperative cleaner fish compared to 50% (50% being the amount of time expected if the behaviour was random (dashed line)). Clients could also approach an unknown cooperation level cleaner. The same experiment was conducted in consecutive years. Median and interquartiles are given.

cleaner's inspection if they witnessed that the cleaner's last interaction ended without conflict, invite less if they do not have such knowledge, and invite the least if the last interaction ended with conflict^{25,26}. Furthermore, cleaner fish are more cooperative in the presence of bystanders than when they are alone with one client²⁶. These results are consistent with the hypothesis that image scoring by client fish is a selfish act used to find a cooperative cleaner or to avoid a cheating cleaner, resulting in (more) cooperative behaviour in cleaners in the presence of bystanders. Here, we test this hypothesis experimentally.

To test whether client fish engage in image-scoring behaviour, we offered a client fish the choice of interacting with two cleaner fish (Fig. 1): one cleaner appeared to behave 'cooperatively' with a model of a client whereas the other did not interact with the model (that is, it had an 'unknown cooperative level'). Client fish spent significantly more than 50% of their time near the cooperative than near the unknown cooperative level cleaner (Fig. 2). This experiment was done twice in consecutive years and both results were significant (one sample Wilcoxon signed rank-sum test with $H_0 = 50\%$: 2002, $n = 14$, $T = 18.5$, $P = 0.04$; 2003, $n = 14$, $T = 16$, $P = 0.02$).

The second experiment, which tested whether cleaner fish foraging behaviour differed in an image-scoring situation compared with a non-image-scoring situation, was based on the following logic: because cleaners may prefer client mucus over ectoparasites¹³, they must feed against their preference (that is, not cheat) during a current interaction to gain access to an observing client. After previous experiments on partner control mechanisms in cleaning mutualism²⁷, we replaced clients, mucus and parasites with plates on levers, and by presenting preferred prawn and less-preferred fish flakes on the plates, respectively (Fig. 3). The levers allowed us to control the plates' response to cleaner fish foraging behaviour according to predetermined rules. As a baseline, all cleaners were offered only one plate (single plate) (Fig. 3a). Furthermore, cleaners were tested in one of two scenarios involving two plates presented simultaneously. In the control no-image-scoring scenario, each plate remained in the aquarium until the cleaner ate one prawn item off it (Fig. 3b). In the experimental image-scoring scenario, the removal of one plate when a cleaner fish fed on prawn led to the immediate removal of the second plate (so the cleaner could not eat anything from the latter) (Fig. 3c). Thus, in the image-scoring scenario, cleaners could eat only the two flake items (FF) on the first plate to then access the second plate, ignoring the two remaining prawn items on the first plate (Fig. 3c). On all other plates, they could eat two flakes and then a prawn item (FFP) for maximal foraging success. Therefore, to maximize their food intake, we predicted that cleaners should eat more against their preference (a ratio of flake items eaten to total items eaten) while interacting with the first plate in the image-scoring scenario than while feeding from any other plate. Comparing the single-plate situation against the two-plate situation tested the

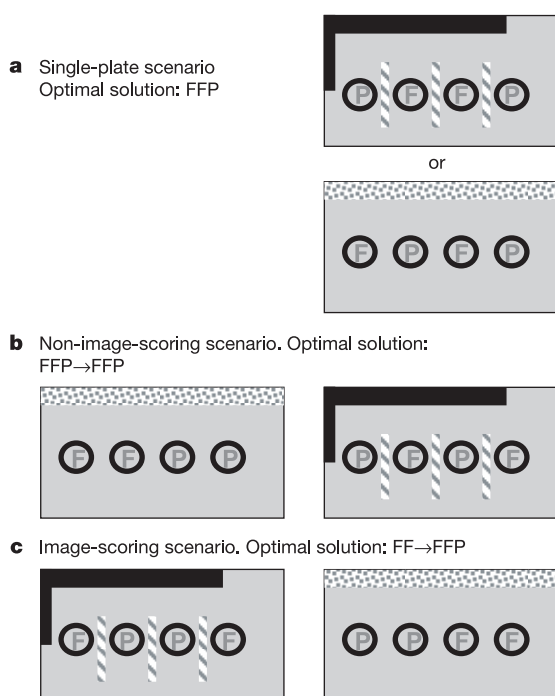


Figure 3 | Experimental set-up to test for indirect reciprocity. Optimal feeding solutions in a foraging experiment where eating fish flakes (F) was allowed whereas eating prawn (P) resulted in immediate termination of an interaction due to plate removal. **a**, Single-plate situation. **b**, Two plates are removed independently when the cleaner fish eats a prawn item. **c**, Two-plate situation where eating prawn on one leads to the removal of both plates. Half of the cleaners were tested in **a** and **b**, the other half in **a** and **c**. Plates were coloured in two ways to allow discrimination: hatched area, pink; stippled area, beige.

alternative hypothesis that cleaner fish may be more cooperative when more food is available, independently of any image-scoring behaviour of clients.

Within treatment groups, we found that in the image-scoring scenario, cleaners fed significantly more against their preference when interacting with the first plate in the two-plate situation than when interacting with the second plate or in the single-plate situation (Friedman test, $n = 12$, $\chi^2_2 = 12.5$, $P = 0.002$, post-hoc multiple comparisons $P < 0.01$) (Fig. 4). In the non-image-scoring scenario, the foraging behaviour of cleaners on all three plates did not differ significantly (Friedman test, $n = 12$, $\chi^2_2 = 1.3$, $P > 0.05$) (Fig. 4). Between treatment groups, cleaners ate significantly more against their preference on the first plate in the image-scoring scenario than on the first plate in the non-image-scoring (control) scenario (Mann–Whitney U -test, $m = 12$, $n = 12$, $Z = 3.15$, $P = 0.002$; Fig. 4). Although cleaners did not always eat items in the sequence that would maximize their food intake, they still also ate against their preference (their preference being prawn²⁷) in all situations (Fig. 4). In the image-scoring scenario, all cleaners sometimes ate only flakes from the first plates (between 2–12 out of 20 trials, median = 7 trials), which allowed them to access the second plate. This happened significantly less often in the no-image-scoring scenario (between 0–4 out of 20 trials, median = 1.5 trials) where the second plate could still be accessed (Mann–Whitney U -test, $m = 12$, $n = 12$, $Z = 3.61$, $P < 0.001$).

Our results show that both requirements for simple indirect reciprocity—image scoring by clients and an increased level of cooperation by cleaners in the presence of image-scoring clients—exist in a cleaning mutualism. Our experimental results confirm interpretations based on field observations^{25,26}. Under natural conditions, cleaners feed against their preference (that is, eat ectoparasites rather than client mucus) to reduce the risk of conflict with their

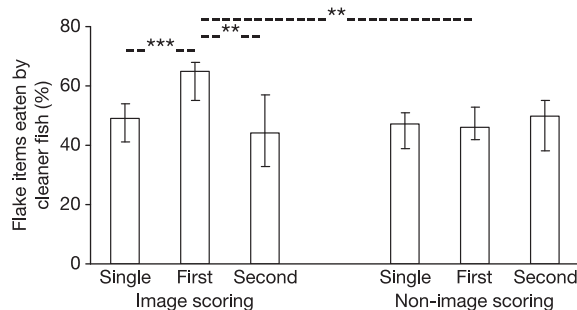


Figure 4 | Indirect reciprocity because of image scoring. Percentage of flake items eaten relative to all items eaten (flake and prawn) in 20 rounds. Twelve fish were tested with single plates and two plates in the image-scoring scenario, and twelve fish were tested with single plates and two plates in the non-image-scoring scenario. Each cleaner contributed one mean value to the single-, first- and second-plate situations. Double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$ in post-hoc comparisons. Median and interquartiles are given.

current client in order to gain access to nearby observing clients²⁵. Although eavesdropping in communication networks may promote increased levels of aggression in the context of animal conflict^{15–17,28}, our data on cleaning mutualism show that eavesdropping may also result in increased levels of cooperation. In our system, indirect reciprocity based on image-scoring clients probably requires only relatively simple cognitive processes. Clients do not have to remember observed interactions as they react immediately to what they observe. Cleaners have over 2,000 client interactions per day¹¹ and therefore constantly receive feedback about the consequences of their actions, allowing them to associate their behaviour with foraging success. It will be a challenge to find more complex forms of indirect reciprocity and image scoring⁵—in particular the purely altruistic system that humans are capable of playing²—in other animals.

METHODS

For additional information, see Supplementary Information

Image-scoring behaviour in client fish. Trials were done in an experimental aquarium with three compartments, separated from each other by one-way mirrors. The mirrors allowed viewing into the side compartments from the middle one (Fig. 1). We placed two client models in the side compartments that each held a cleaner fish, *Labroides dimidiatus*. To create differences between the inspection behaviour of the two cleaner fish with their model, only one model had mashed prawn (0.004 g) spread on it. The model with food was placed in the compartment holding cleaner fish trained to feed off the model, whereas the model without food was placed in the compartment holding cleaner fish familiar to a model having no food on it. Therefore, the former cleaner would interact more with the model (cooperative cleaner) than the latter (unknown cooperation level cleaner). This method worked, as 27 out of 28 cleaners with a model with prawn on it interacted longer with the model than their 'partner' cleaner on the other side. Cooperative cleaners interacted on average for 250 s with the model (range 20–550 s), about 12.5 times longer than cleaners with a model without food (average 20 s, range 0–130 s). To control for the presence of food, an equal amount of prawn was placed in the other cleaner's compartment, but it dropped to the bottom where it was ignored by the cleaners. Food on the model was on the side facing the client *Scolopsis bilineatus* in the middle compartment. Opaque barriers covering the mirrors were removed as soon as both models were in place and fish behaviour recorded for 10 min. We recorded the amount of time the client spent in a designated area near each cleaner (Fig. 1) and the time each cleaner spent interacting with the model, estimated by a scan sample at 10-s intervals. Trials were scheduled so that cooperative cleaners were at each side of the aquaria in 50% of trials to account for any preference of the client for a particular side. Clients spent on average 320 s (range 86–599 s) of their time near either cleaner fish (that is, with their full body out of the shelter to either side).

Cleaner fish indirect reciprocity experiment. The experimental set-up allowed us (1) to offer two types of food as distinct items visible to the experimenter, hence enabling us to see exactly what item a cleaner fish ate with each bite; (2) to remove a plate immediately when a cleaner ate a preferred food item; and (3) to

offer two plates simultaneously and manipulate the presence of the second plate according to the cleaner's foraging behaviour.

First learning phase. Following methods described in ref. 13, we trained cleaners that eating less-preferred fish flakes was allowed whereas eating preferred prawn led to the immediate removal of a plate. Plate removal was done using a lever attached to the plate. A total of six learning trials was done every 2 h, four trials per day.

Second learning phase and experiment. We conducted six rounds of the experiment without scoring the cleaners' behaviour, followed by 20 experimental rounds. We used two different-looking plates (10 × 6 cm) (Fig. 3). Each round consisted of two trials 30 min apart, one with two plates present and one with a single plate present. The following round started 60 min later, so the experiment was spread over 3 days. Twelve cleaners were used for each of the control and experimental groups. The key difference between the experimental and the control group occurred in the two-plate situation: the control cleaners (non-image-scoring scenario) could eat from the first plate until they ate a prawn item, after which the plate was removed, and then move on to the second plate until they ate a prawn item, after which the second plate was removed (Fig. 3b). Thus, the second bystander plate did not form an image of the cleaner's behaviour. The experimental (image-scoring scenario) cleaners, however, faced the problem that eating prawn on one plate led to the removal of both plates (Fig. 3c). Thus, the second bystander plate could form an image and respond negatively to the cleaner, by leaving and avoiding an interaction, if the cleaner cheated the first plate. Given these rules of plate behaviour, cleaner fish had to use varying rules to optimize their foraging success: the optimal solutions shown in Fig. 3.

The cleaner was kept to one half of the aquarium with a clear partition, while the plate(s) was placed against an aquarium side (always the same side) in the other half. Once the plate(s) was in place, we removed the partition so that the cleaner could begin foraging. For each cleaner, we calculated the percentage of flake items eaten per plate per scenario.

Received 31 January; accepted 22 March 2006.

- Alexander, R. D. *The Biology of Moral Systems* (Aldine de Gruyter, New York, 1987).
- Wedekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **288**, 850–852 (2000).
- Fehr, E. & Gächter, U. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
- Milinski, M., Semmann, D. & Krambeck, H. J. Reputation helps solve the 'tragedy of the commons'. *Nature* **415**, 424–426 (2002).
- Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity. *Nature* **437**, 1291–1298 (2005).
- Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577 (1998).
- Leimar, O. & Hammerstein, P. Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B* **268**, 745–753 (2001).
- Zahavi, A. Altruism as a handicap—the limitations of kin selection and reciprocity. *J. Avian Biol.* **26**, 1–3 (1995).
- Roberts, G. Competitive altruism: from reciprocity to the handicap principle. *Proc. R. Soc. Lond. B* **265**, 427–431 (1998).
- Lotem, A., Fishman, M. A. & Stone, L. From reciprocity to unconditional altruism through signaling benefits. *Proc. R. Soc. Lond. B* **270**, 199–205 (2003).
- Grutter, A. S. Spatio-temporal variation and feeding selectivity in the diet of the cleaner fish *Labroides dimidiatus*. *Copeia* **1997**, 346–355 (1997).
- Arnal, C. & Côté, I. M. Diet of broadstrip cleaning gobies on a Barbadian reef. *J. Fish Biol.* **57**, 1075–1082 (2000).
- Grutter, A. S. & Bshary, R. Cleaner wrasse prefer client mucus: support for partner control mechanisms in cleaning interactions. *Proc. R. Soc. Lond. B* **270** (suppl.), 242–244 (2003).
- McGregor, P. K. Signalling in territorial systems: a context for individual identification, ranging and eavesdropping. *Phil. Trans. R. Soc. Lond. B* **340**, 237–244 (1993).
- McGregor, P. K. (ed.) *Animal Communication Networks* (Cambridge Univ. Press, Cambridge, 2005).
- Johnstone, R. A. Eavesdropping and animal conflict. *Proc. Natl Am. Soc.* **98**, 9177–9180 (2001).
- Johnstone, R. A. & Bshary, R. The evolution of spiteful behaviour. *Proc. R. Soc. Lond. B* **271**, 1917–1922 (2004).
- Sherratt, T. N. & Roberts, G. The role of phenotypic defectors in stabilizing reciprocal altruism. *Behav. Ecol.* **12**, 313–317 (2001).
- Matos, R. J. & Schlupp, I. In *Animal Communication Networks* (ed. McGregor, P. K.) 84–113 (Cambridge Univ. Press, Cambridge, 2005).
- Bshary, R. & Bronstein, J. L. Game structures in mutualisms: what can the evidence tell us about the kind of models we need? *Adv. Stud. Behav.* **34**, 59–101 (2004).
- Bshary, R. & Noë, R. In *Genetic and Cultural Evolution of Cooperation* (ed. Hammerstein, P.) 167–184 (MIT Press, Cambridge, 2003).
- Trivers, R. L. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).

23. Bshary, R. In *Economics in Nature* (eds Noë, R., van Hooff, J. A. R. A. M. & Hammerstein, P.) 146–172 (Cambridge Univ. Press, Cambridge, 2001).
24. Dugatkin, L. A. *Cooperation Among Animals: An Evolutionary Perspective* (Oxford Univ. Press, Oxford, 1997).
25. Bshary, R. Biting cleaner fish use altruism to deceive image scoring clients. *Proc. R. Soc. Lond. B* **269**, 2087–2093 (2002).
26. Bshary, R. & D'Souza, A. In *Communication Networks* (ed. McGregor, P.) 521–539 (Cambridge Univ. Press, Cambridge, 2005).
27. Bshary, R. & Grutter, A. S. Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biol. Lett.* **1**, 396–399 (2005).
28. Dutreland, C., McGregor, P. K. & Oliveira, R. F. The effect of an audience on intrasexual communication in male Siamese fighting fish, *Betta splendens*. *Behav. Ecol.* **12**, 283–286 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Lizard Island Research Station for their continuous support and friendship, and W. Wickler for his support and for discussions on this topic. R. Bergmüller, K. Cheney, P. Munday and M. Blows provided substantial comments on earlier drafts. Funding was provided by NERC (R.B.) and the Australian Research Council (A.S.G.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.B. (redouan.bshary@unine.ch).

Experience-dependent and cell-type-specific spine growth in the neocortex

Anthony Holtmaat¹, Linda Wilbrecht¹, Graham W. Knott², Egbert Welker² & Karel Svoboda¹

Functional circuits in the adult neocortex adjust to novel sensory experience, but the underlying synaptic mechanisms remain unknown¹. Growth and retraction of dendritic spines with synapse formation and elimination could change brain circuits^{2–7}. In the apical tufts of layer 5B (L5B) pyramidal neurons in the mouse barrel cortex, a subset of dendritic spines appear and disappear over days, whereas most spines are persistent for months^{4–6,8,9}. Under baseline conditions, new spines are mostly transient and rarely survive for more than a week. Transient spines tend to be small^{4,5,9}, whereas persistent spines are usually large^{4–6,8,9}. Because most excitatory synapses in the cortex occur on spines, and because synapse size¹⁰ and the number of α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) receptors^{11–13} are proportional to spine volume, the excitation of pyramidal neurons is probably driven through synapses on persistent spines. Here we test whether the generation and loss of persistent spines are enhanced by novel sensory experience. We repeatedly imaged dendritic spines for one month after trimming alternate whiskers, a paradigm that induces adaptive functional changes in neocortical circuits^{14,15}. Whisker trimming stabilized new spines and destabilized previously persistent spines. New-persistent spines always formed synapses. They were preferentially added on L5B neurons with complex apical tufts rather than simple tufts. Our data indicate that novel sensory experience drives the stabilization of new spines on subclasses of cortical neurons. These synaptic changes probably underlie experience-dependent remodelling of specific neocortical circuits.

With the use of two-photon laser-scanning microscopy^{16,17}, we imaged spines in neocortical layer 1 (L1) through a small glass window, implanted above the barrel cortex of adult transgenic mice that express enhanced green fluorescent protein (eGFP) in a sparse subset of pyramidal neurons (line GFP-M)^{4,5,18} (Fig. 1a, b, d). To measure the dynamics of dendritic spines over times relevant for barrel cortex plasticity¹⁹, we acquired images every four days for at least 28 days (Fig. 1c). We scored clear instances of spine appearance or disappearance, keeping track of spine lifetimes (Fig. 1d, e). The data reported here are from L5B pyramidal neurons with somata in barrel cortex.

Under control conditions, the density of spines ($240 \pm 30 \text{ mm}^{-1}$; control, 6 mice, $n = 7$ cells, 1,592 spines) did not change significantly over the time course of the experiment (Fig. 1f) ($P = 0.08$; Wilcoxon, comparing days 0 and 28). Consistent with previous measurements, a subpopulation of ‘transient’ spines appeared and disappeared (fractional turnover between imaging sessions, 0.23 ± 0.08)^{4,5} with a time-constant of a few days ($\tau \approx 4.6$ days), whereas the majority of spines persisted (‘always-present’ fraction, 0.62 ± 0.08) (Supplementary Fig. 1).

We focused our analysis on persistent spines, defined as spines that lived for eight days or more (at least three imaging sessions), double

the time constant of transient spines. ‘New-persistent’ spines (NP) were defined as spines that were gained and remained present through the last eight days of imaging (see Fig. 1c for the definition of spine categories). ‘Lost-persistent’ spines (LP) were defined as spines present during the first eight days of imaging that subsequently disappeared. The number of NP and LP spines was a small fraction of the total number of persistent spines (Fig. 1g, i): new-persistent (NP/(AP+NP)), 0.05 ± 0.04 ; lost-persistent (LP/(AP+LP)), 0.10 ± 0.05 (AP, number of always-present spines). These measurements confirm that under control conditions the addition and subtraction of persistent spines are rare^{4–6,8,9}.

Do changes in experience drive gains and losses of persistent spines? In the adult barrel cortex, trimming of a subset of whiskers induces functional reorganization of intracortical circuits^{14,15}, including circuits impinging on L5B neurons (N. Wright and K. Fox, personal communication). To test whether whisker trimming can affect persistent spines, we cut the contralateral whiskers such that each deprived cortical barrel column was surrounded by four spared barrel columns (Fig. 1b)^{4,15}. This ‘chessboard’ trimming followed an eight-day baseline period and lasted for at least 20 days (Fig. 1c) (trimmed, ten mice, $n = 11$ cells, 3,933 spines). The baseline period allowed us to identify persistent spines and to assess the impact of changing experience on their stability. The generation of new-persistent spines was measured from the day of trimming (day 8, Fig. 1c).

After trimming, the density of spines ($240 \pm 70 \text{ mm}^{-1}$) did not change detectably (Fig. 1f) ($P = 0.46$; Wilcoxon, comparing days 0 and 28). The total number of new spines (TN, spines first detected on days 8–20; Fig. 1c), which is dominated by the appearance of transient spines, also remained unchanged by whisker trimming ($P = 0.40$).

Whisker trimming enhanced the fraction of NP spines (Fig. 1e, orange arrowheads) compared with controls (Fig. 1g) (fraction NP; control, 0.05 ± 0.04 ; trimmed, 0.13 ± 0.07 , $P < 0.001$). Because spine densities varied more than twofold from neuron to neuron (Fig. 1f), we also analysed the density of NP spines. Whisker trimming increased the density of NP spines (Fig. 1h) (NP density; control, $8 \pm 7 \text{ mm}^{-1}$; trimmed, $23 \pm 15 \text{ mm}^{-1}$; $P < 0.005$).

A disproportionately large fraction of the spine gains after whisker trimming were on dendrites located in regions between barrel columns (‘septum-related columns’²⁰), in comparison with dendrites in deprived or spared barrel columns (NP density: septum, $30 \pm 4 \text{ mm}^{-1}$; spared barrels, $17 \pm 3 \text{ mm}^{-1}$; deprived barrels, $16 \pm 4 \text{ mm}^{-1}$; septum versus barrels, $P < 0.01$) (Supplementary Information and Supplementary Fig. 2). Distinct classes of thalamocortical and intracortical afferents are concentrated in barrel-related and septum-related columns^{20,21}. Different branches of individual neurons might therefore, by virtue of their location, participate in distinct circuits and distinct forms of plasticity.

¹Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. ²Institut de biologie cellulaire et de morphologie (IBCM), Université de Lausanne, Rue du Bugnon 9, CH 1005 Lausanne, Switzerland.

We also analysed the effect of trimming whiskers on the loss of persistent spines (Fig. 1e, green arrowheads). Under control conditions, persistent spines disappeared at somewhat higher rates than they appeared^{5,6}, although the difference was not significant ($P = 0.06$). After whisker trimming, persistent spines were more likely to disappear (Fig. 1i, j) (fraction LP: control, 0.10 ± 0.05 ; Trimmed, 0.15 ± 0.03 , $P < 0.01$; LP density: control, $17 \pm 9 \text{ mm}^{-1}$; trimmed, $25 \pm 7 \text{ mm}^{-1}$, $P < 0.05$). Our results show that changes in sensory input, induced by chessboard whisker trimming, enhance the addition and subtraction of persistent spines, keeping total spine numbers constant (for a comparison with other sensory-deprivation models see Supplementary Information).

Previous work has shown that some new spines^{4,22} and axonal boutons⁷ have synapses, indicating that structural plasticity might underlie functional changes in brain circuits. To determine whether NP spines have synapses, we reconstructed a subset of NP spines (11 spines, 3 mice) by retrospective serial section electron microscopy⁴ (Fig. 2a–d). Spines were considered to have synapses only if they had a postsynaptic density (PSD) and an active zone directly apposed to the

PSD, containing synaptic vesicles. NP spines (at least eight days old) always had synapses (11/11; Fig. 2b, d). This indicates that the addition of NP spines is consistent with synapse formation²³. The addition of persistent spines might thus underlie functional changes in excitatory circuits impinging on L5B pyramidal cells.

The relationship between transient spines and NP spines is unclear. New transient spines may serve to generate large numbers of diverse, short-lived contacts. Changes in sensory experience could lead to the selection of a subset of these connections for stabilization (sample-and-select model)²⁴. This model makes two testable predictions. First, whisker trimming should enhance the probability that new spines become persistent. We counted all new spines that were first detected on the day of trimming or over the following 12 days (TN) and scored the subpopulation of these spines that survived to the end of the experiment and thus turned into NP spines (Fig. 1c). The probability that new spines became persistent was enhanced after trimming compared with control conditions; this is consistent with the sample-and-select model (Fig. 2e) (probability stabilized, NP/TN: control, 0.04 ± 0.02 ; trimmed, 0.09 ± 0.04 ; $P < 0.001$).

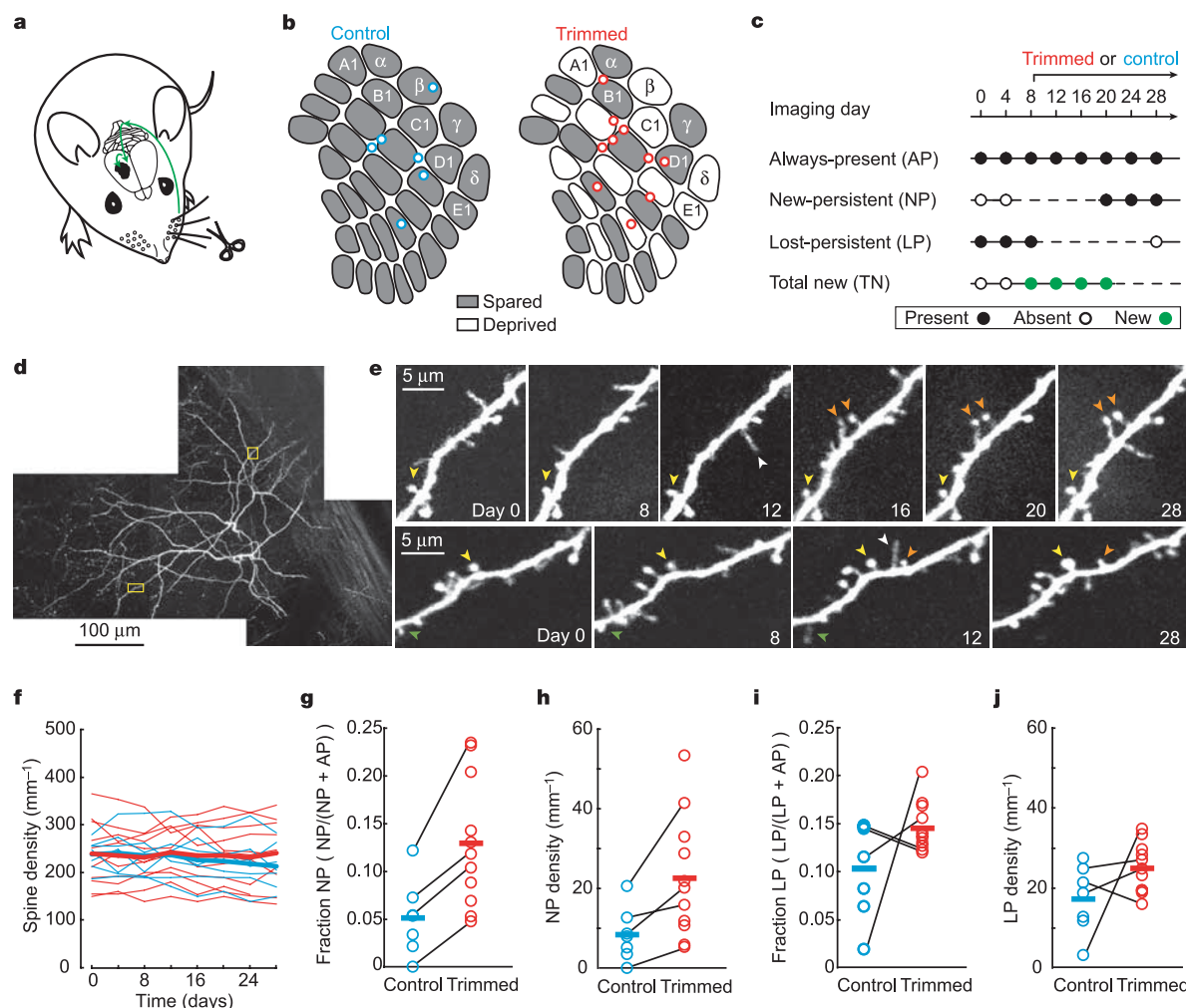


Figure 1 | Retraction and growth of persistent spines. **a**, Diagram of the whisker-to-barrel pathway and chessboard whisker trimming. **b**, Locations of the somata of imaged neurons in the barrel map (blue, control; red, trimmed). Two somata, one in each group, were not precisely mapped. **c**, Timeline of the experiment. In a subset of experiments, whiskers were trimmed immediately after the third imaging session. Definitions of spine categories used in the analysis are as follows: AP spines were present for all 28 days; NP spines were first observed on the day of trimming or after the baseline period, and were present for the last eight days of imaging; LP spines were present during the eight days of baseline period and were subsequently

lost; TN spines include all new spines appearing on days 8–20. **d**, Overview of the apical tuft of an L5B neuron. **e**, High-magnification view of two dendritic segments imaged before (days 0–8) and after trimming (days 12–28) (from boxed regions in **d**). Arrowheads: yellow, AP; orange, NP; green, LP; white, transient. **f**, Spine density. **g**, **h**, Quantification of LP and NP spines. Every circle represents a cell. Horizontal bars represent the group averages. Black lines connect paired experiments, in which a complete control period was followed by 20 days of deprivation. **g**, Fraction of NP spines, NP/(AP+NP). **h**, Density of NP spines. **i**, Fraction of LP spines, LP/(AP+LP). **j**, Density of LP spines.

Second, because persistent spines tend to be larger than transient spines⁵, stabilization of transient spines should be accompanied by an increase in spine size. We therefore analysed structural changes of NP spines after their emergence. NP spines sometimes appeared as faint, thin spines and in later imaging sessions acquired a bright spine head (Fig. 2f, orange arrowheads; Supplementary Fig. 3). We quantified these changes by measuring spine brightness, which is monotonically related to spine volume⁵ (Fig. 2g–i). The brightness of NP

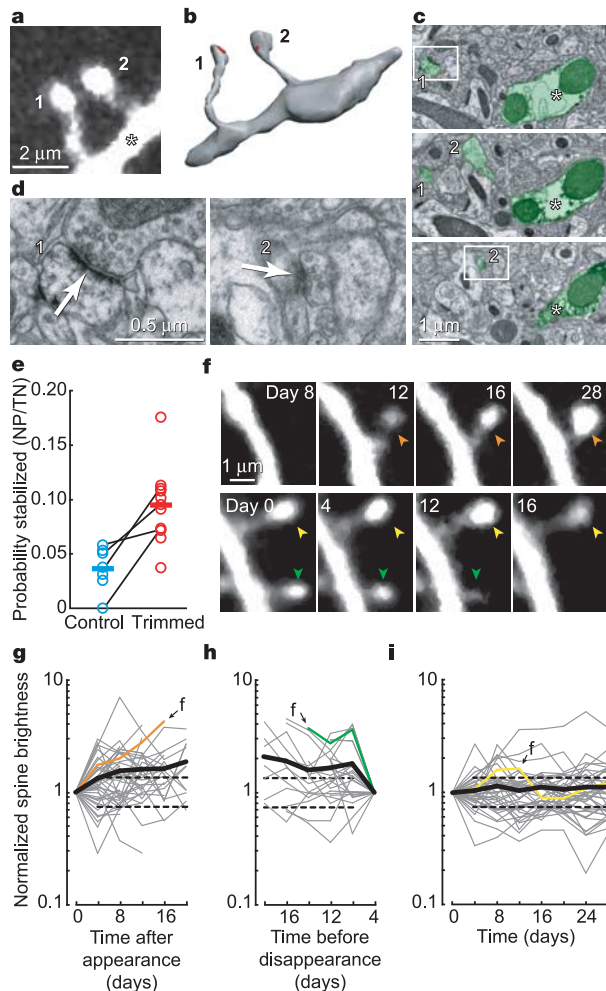


Figure 2 | Spine stabilization and destabilization. **a**, *In vivo* image of two NP dendritic spines that were subsequently reconstructed with serial-section electron microscopy (EM). **b**, Serial-section EM reconstruction of spines from **a**. Synapses are drawn in red. Spine 2 has two synapses, only one of which is visible. **c**, EM images of individual serial sections (examples are not consecutive). The parent dendrite (asterisk) and spine heads 1 and 2 are coloured green. **d**, High-magnification micrographs through spines 1 (left) and 2 (right). Arrows indicate synaptic densities (synapse 2 is not visible in **b**). **e**, The probability (NP/TN) that new spines become persistent. **f**, Change in brightness of spines after their appearance (orange arrowheads) and before their loss (green arrowheads) and that of an AP spine (yellow arrowheads). In this example, the new spine appeared between day 8 and 12 and subsequently became brighter and thus larger. The lost spine became dimmer and thus smaller before it disappeared some time between days 12 and 16. **g**, Normalized brightness of NP spines aligned (and normalized) to the day of their appearance ($t = 0$) (coloured line, example from **f**; grey lines, individual spines; black line, average spine brightness; dashed lines, estimate of the maximum apparent changes in volume due to movement noise during imaging; see Methods). **h**, Normalized brightness of LP spines aligned (and normalized) to the imaging day before their disappearance ($t = 4$ days). **i**, Normalized brightness of AP spines aligned (and normalized) to the first imaging day ($t = 0$).

spines increased significantly after they appeared (Fig. 2g) (average normalized change in brightness, 1.47 ± 0.64 ; 43 spines; $P < 0.001$, Wilcoxon, comparing first time point with the average of all subsequent time points). This indicates that spines continue to grow after their appearance, which is consistent with the sample-and-select model. Conversely, spines that were destined to disappear often decreased in brightness just before their disappearance (Fig. 2f, green arrowheads; Fig. 2h) (average normalized change in brightness, 0.73 ± 0.38 ; 25 spines; $P < 0.01$, Wilcoxon, comparing the last time point with the average of the previous time points). In contrast, the brightness of AP spines remained unchanged on average, although individual AP spines underwent large increases and decreases in brightness (Fig. 2f, yellow arrowheads; Fig. 2i) (41 spines; $P = 0.99$; Wilcoxon, comparing day 8 with the average of all subsequent time points). Changes in spine volume might reflect increases and decreases in synaptic strength^{10–13}, or they might simply signal synapse formation and elimination.

The fraction of NP and LP spines varied greatly from neuron to neuron (Fig. 1g, i) ($P < 0.001$; comparison of NP densities assuming a Poisson distribution of spines). To understand the basis of this heterogeneity we analysed the morphology of imaged neurons (Supplementary Information; $n = 14$). Although all imaged neurons shared the same cortical layer, the structure of their apical tufts varied. At one end of the spectrum, cells branched close to L1 with few apical branches. They had small total dendritic lengths, low maximum branch orders and low spine densities (Fig. 3a, and Supplementary Fig. 4). At the other end of the spectrum, cells had a major dendrite bifurcation in deeper layers of the cortex (L4–L2). Their apical tufts had numerous apical branches, large total lengths, high maximum branch orders and high spine densities (Fig. 3b, and Supplementary Fig. 4). All of these structural parameters were highly correlated (Fig. 3c, and Supplementary Fig. 4). The fraction and density of NP spines increased with increasing complexity (Fig. 3d, e; $P < 0.02$). In contrast, the fraction and density of LP spines were independent of complexity (Fig. 3f, g; $P = 0.65$).

On the basis of dendritic morphology, cells could be separated into two groups, composed of simple-tuft and complex-tuft cells (k -means clustering; differences between groups, $P < 0.005$; Fig. 3c, and Supplementary Fig. 4). Their anatomical characteristics correspond to regular spiking (RS) and intrinsically bursting (IB) neurons in L5B described previously^{25–28}. Consistent with the correlations between spine growth and dendritic complexity (Fig. 3d, e) was the observation that after trimming, NP spines were more likely to appear on complex-tuft cells than on simple-tuft cells (see Supplementary Information for details). In contrast, losses of persistent spines occurred with similar frequencies in both cell types. The net effect of these changes was a decrease in the density of persistent spines in simple-tuft cells and an increase in the density of persistent spines in complex-tuft cells after whisker trimming (final density divided by initial density; simple-tuft cells, 0.92 ± 0.09 ; complex-tuft cells, 1.07 ± 0.07 ; $P < 0.01$). Thus the experience-dependent spine plasticity of cortical neurons depends on the cell type, even for neurons sharing the same cortical sublayer.

At the level of circuits, intracortical excitatory projections onto distinct types of L5B neurons²⁵ reorganize differently in response to trimming: RS (simple; small apical tufts) cells lose intracortical input, whereas IB (complex; large apical tufts) cells gain intracortical input (L. Petreanu, G. Shepherd and K.S., unpublished data). These findings are consistent with our data because we found that persistent spine density decreased on simple-tuft (RS) cells and increased on complex-tuft (IB) cells (Fig. 3d–g; see also Supplementary Information). The alignment of these observations strengthens the evidence that new spines contribute to experience-dependent changes in excitatory circuits, either through the formation of synapses between previously unconnected neurons or by increasing the number of synapses between previously connected neurons.

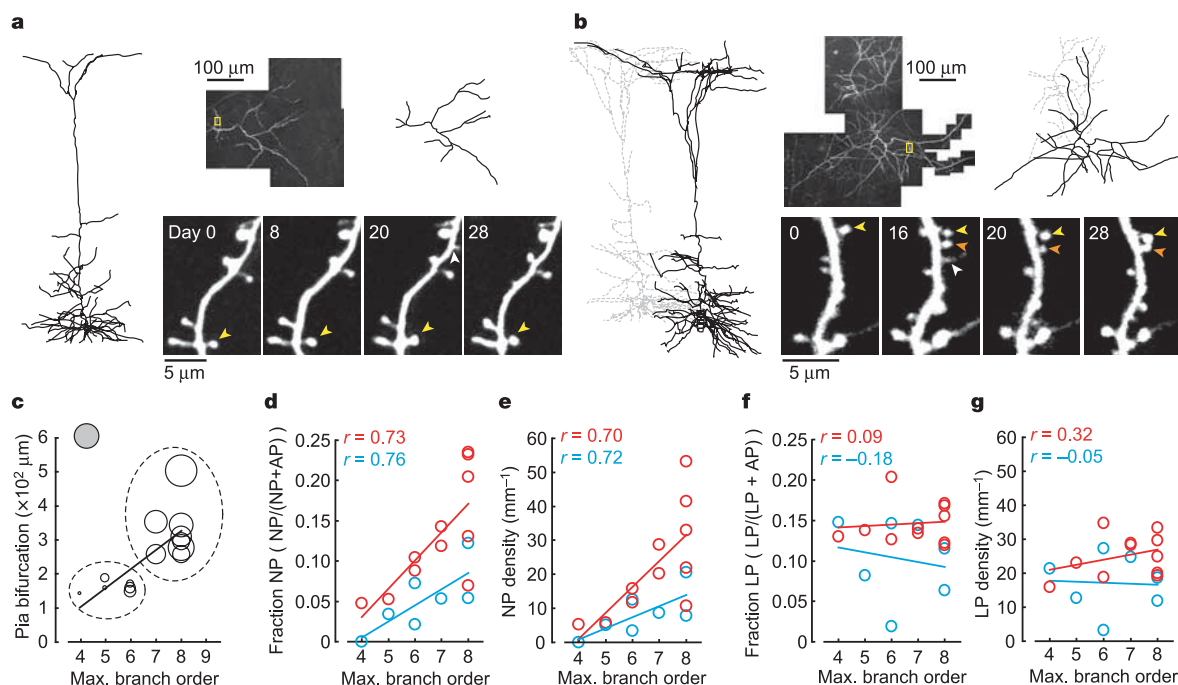


Figure 3 | Cell-type-dependent growth of persistent spines. **a, b**, Examples of imaged pyramidal cells in L5B. **a**, Simple-tuft cell (left, side view; right, top view; bottom, high-magnification time-lapse of boxed area). Arrowheads: yellow, AP; orange, NP; white, transient. **b**, Two complex-tuft cells; arrangement as in **a**. **c**, Correlations between selected structural parameters. Three of the five parameters used to distinguish simple-tuft from complex-

tuft cells are plotted. The two distinct cell clusters are encircled by dashed lines. The grey scale circle at the top left represents a total branch length of 5 mm. **d–g**, Persistent spine changes as a function of apical tuft complexity (maximum branch order). Each circle represents a cell. Blue, control; red, trimmed. Correlation coefficients (r) are shown. **d**, Fraction of NP spines. **e**, Density of NP spines. **f**, Fraction of LP spines. **g**, Density of LP spines.

METHODS

Surgery and imaging. Male C57BL/6J transgenic mice (ages 2–5 months; median age ~ 2.3 months) expressing eGFP under control of the Thy-1 promoter (line M)¹⁸ were used for this study (13 mice, 14 cells). Surgeries and imaging were performed as described^{5,29} (Supplementary Methods).

To change the tactile input from the facial vibrissae to the barrel cortex, we trimmed whiskers on the contralateral mystacial pad in such a way that each trimmed whisker was surrounded by intact whiskers (chessboard deprivation). The whiskers on the ipsilateral mystacial pad were all trimmed. Whisker trimming was performed every two days under brief anaesthesia with isoflurane, and was started after a baseline imaging period of at least eight days, after the third imaging session (Fig. 1a–c). For control mice, mock trimming was performed under identical conditions by sweeping the whiskers back and forth without damaging them.

Four cells (three mice) were monitored for a 28-day control period followed by a 20-day deprivation period. These cells are illustrated in Fig. 1g–j as paired values connected by black lines. In total, 7 cells were imaged over a control period and 11 cells over a trimmed period (Fig. 1f–j).

Image analysis. We analysed a total of about 7 mm of dendrite over 8–13 imaging sessions, containing 1,794 dendritic spines at the start of the imaging period, and a total of 5,333 spines over the entire imaging period (including spines that were gained and lost). All analysis was done in three dimensions with custom software. Only spines emanating laterally from the dendritic shaft for more than five pixels (0.4 μ m) were measured. Because even thin spines can form synapses^{4,10,30}, we included all spines in our measurements irrespective of their shapes, including structures that others have classified as ‘filopodia’^{6,8}. All spine counts were done with the analyser being unaware of the experimental condition.

For analysis of spine brightness, spines were initially selected from a database without reference to the images. Subsequently, some of these spines were rejected for analysis if they were not clearly separated from their parent dendrites. We analysed 41 APs, 43 NPs and 25 LPs, from 22 dendritic segments and 7 cells, in their best focal section. After background subtraction, the pixel values containing the spine head were summed. Throughout the text we refer to ‘spine brightness’ as the spine fluorescence divided by the fluorescence of the local dendrite. Because dendritic shaft diameters were constant and relatively uniform, we used them to correct spine fluorescence levels for variations in excitation levels over time. For each spine we estimated the fluorescence of a

segment of its parent dendrite (length 5–15 μ m) by averaging the pixel values along its medial axis. To estimate the apparent spine brightness changes due to movement noise, we imaged 26 spines on three dendritic segments with short intervals (1–5 min). The ratio of the highest to the lowest spine brightness over four time points was averaged over all spines and served as an estimate of the maximum change of spine brightness due to movement artefacts. This noise baseline is indicated in the Fig. 2g–i as dashed lines.

Anatomy. Serial-section electron microscopy, with a pre-embedding immunolabelling protocol, was performed as described⁴. For all other retrospective reconstructions and localization of imaged cells, the mice were perfused with 4% paraformaldehyde and immunostained⁴. The barrel walls in layer 4 were revealed with Nissl staining (Supplementary Information and Supplementary Fig. 2). The blood vessels in each section were used for alignment and to locate the imaged cell, using the blood vessel maps from images *in vivo* as a reference. The entire apical dendritic tuft, stalk and the soma of each cell were reconstructed with Neurolucida software (MicroBrightfield). All somata were in the barrel field ($n = 14$). Their precise locations were reconstructed relative to individual barrels for all except two cells (these are lacking in Fig. 1b). For analysis of the cell morphology we used neurolucida reconstructions from stained sections as well as from images *in vivo* ($n = 14$). Partitioning of neurons based on apical tuft complexity was performed by using a k -means cluster analysis (Supplementary Information and Supplementary Fig. 4).

Statistics. We used non-parametric bootstrap methods to compute differences between means and medians and report the larger P value. Paired comparisons of longitudinal measurements used the Wilcoxon signed rank test as indicated in the text. Significance was set at $P = 0.02$. Parameter values in the text are means $\pm$ s.d.

Received 25 January; accepted 4 April 2006.

1. Buonomano, D. V. & Merzenich, M. M. Cortical plasticity: from synapses to maps. *Annu. Rev. Neurosci.* **21**, 149–186 (1998).
2. Bailey, C. H. & Kandel, E. R. Structural changes accompanying memory formation. *Annu. Rev. Physiol.* **55**, 397–426 (1993).
3. Knott, G. W., Quairiaux, C., Genoud, C. & Welker, E. Formation of dendritic spines with GABAergic synapses induced by whisker stimulation in adult mice. *Neuron* **34**, 265–273 (2002).
4. Trachtenberg, J. T. *et al.* Long-term *in vivo* imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* **420**, 788–794 (2002).

5. Holtmaat, A. J. *et al.* Transient and persistent dendritic spines in the neocortex *in vivo*. *Neuron* **45**, 279–291 (2005).
6. Zuo, Y., Yang, G., Kwon, E. & Gan, W. B. Long-term sensory deprivation prevents dendritic spine loss in primary somatosensory cortex. *Nature* **436**, 261–265 (2005).
7. De Paola, V. *et al.* Cell type-specific structural plasticity of axonal branches and boutons in the adult neocortex. *Neuron* **49**, 861–875 (2006).
8. Grutzendler, J., Kasthuri, N. & Gan, W. B. Long-term dendritic spine stability in the adult cortex. *Nature* **420**, 812–816 (2002).
9. Majewska, A. K., Newton, J. R. & Sur, M. Remodeling of synaptic structure in sensory cortical areas *in vivo*. *J. Neurosci.* **26**, 3021–3029 (2006).
10. Harris, K. M. & Stevens, J. K. Dendritic spines of CA1 pyramidal cells in the rat hippocampus: serial electron microscopy with reference to their biophysical characteristics. *J. Neurosci.* **9**, 2982–2997 (1989).
11. Nusser, Z. *et al.* Cell type and pathway dependence of synaptic AMPA receptor number and variability in the hippocampus. *Neuron* **21**, 545–559 (1998).
12. Takumi, Y., Ramirez-Leon, V., Laake, P., Rinovik, E. & Ottersen, O. P. Different modes of expression of AMPA and NMDA receptors in hippocampal synapses. *Nature Neurosci.* **2**, 618–624 (1999).
13. Matsuzaki, M. *et al.* Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons. *Nature Neurosci.* **4**, 1086–1092 (2001).
14. Diamond, M. E., Huang, W. & Ebner, F. F. Laminar comparison of somatosensory cortical plasticity. *Science* **265**, 1885–1888 (1994).
15. Fox, K. Anatomical pathways and molecular mechanisms for plasticity in the barrel cortex. *Neuroscience* **111**, 799–814 (2002).
16. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning microscopy. *Science* **248**, 73–76 (1990).
17. Denk, W. & Svoboda, K. Photon upmanship: why multiphoton imaging is more than a gimmick. *Neuron* **18**, 351–357 (1997).
18. Feng, G. *et al.* Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41–51 (2000).
19. Glazewski, S. & Fox, K. Time course of experience-dependent synaptic potentiation and depression in barrel cortex of adolescent rats. *J. Neurophysiol.* **75**, 1714–1729 (1996).
20. Shepherd, G. M., Polgruto, T. A. & Svoboda, K. Circuit analysis of experience-dependent plasticity in the developing rat barrel cortex. *Neuron* **38**, 277–289 (2003).
21. Lu, S. M. & Lin, R. C. S. Thalamic afferents of the rat barrel cortex: a light- and electron-microscopic study using *Phaseolus vulgaris* leucoagglutinin as an anterograde tracer. *Somatosens. Mot. Res.* **10**, 1–16 (1993).
22. Toni, N., Buchs, P. A., Nikonenko, I., Bron, C. R. & Muller, D. LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite. *Nature* **402**, 421–425 (1999).
23. Yuste, R. & Bonhoeffer, T. Genesis of dendritic spines: insights from ultrastructural and imaging studies. *Nature Rev. Neurosci.* **5**, 24–34 (2004).
24. Jontes, J. D. & Smith, S. J. Filopodia, spines, and the generation of synaptic diversity. *Neuron* **27**, 11–14 (2000).
25. Chagnac-Amitai, Y., Luhmann, H. J. & Prince, D. A. Burst generating and regular spiking layer 5 pyramidal neurons of rat neocortex have different morphological features. *J. Comp. Neurol.* **296**, 598–613 (1990).
26. Larkman, A. & Mason, A. Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex. I. Establishment of cell classes. *J. Neurosci.* **10**, 1407–1414 (1990).
27. Kasper, E. M., Larkman, A. U., Lubke, J. & Blakemore, C. Pyramidal neurons in layer 5 of the rat visual cortex. I. Correlation among cell morphology, intrinsic electrophysiological properties, and axon targets. *J. Comp. Neurol.* **339**, 459–474 (1994).
28. Tsiola, A., Hamzei-Sichani, F., Peterlin, Z. & Yuste, R. Quantitative morphologic classification of layer 5 neurons from mouse primary visual cortex. *J. Comp. Neurol.* **461**, 415–428 (2003).
29. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: Flexible software for operating laser-scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
30. Fiala, J. C., Feinberg, M., Popov, V. & Harris, K. M. Synaptogenesis via dendritic filopodia in developing hippocampal area CA1. *J. Neurosci.* **18**, 8900–8911 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature

Acknowledgements We thank L. Petreanu for suggesting the comparison of cell types; V. DePaola, B. Burbach and C. Musetti for help with experiments; K. Masback for data analysis; T. O'Connor for writing the spine analysis software; and V. DePaola, G. Shepherd, J. Trachtenberg and K. Zito for comments on the manuscript. This work was supported by the Howard Hughes Medical Institute, the NIH and the Swiss National Science Foundation (G.W.K., E.W.).

Author Contributions A.H. and L.W. contributed equally to this work. A.H. and L.W. performed the imaging experiments. K.S. built the custom microscope. G.W.K. performed the ssEM reconstructions with help from A.H. and L.W. G.W.K., A.H. and L.W. performed the cell reconstructions. A.H., L.W. and K.S. analysed the data and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.S. (svoboda@cshl.edu).

LETTERS

Mammalian cochlear supporting cells can divide and trans-differentiate into hair cells

Patricia M. White^{1*}, Angelika Doetzlhofer^{1*}, Yun Shain Lee¹, Andrew K. Groves^{1,2} & Neil Segil^{1,2}

Sensory hair cells of the mammalian organ of Corti in the inner ear do not regenerate when lost as a consequence of injury, disease, or age-related deafness. This contrasts with other vertebrates such as birds, where the death of hair cells causes surrounding supporting cells to re-enter the cell cycle and give rise to both new hair cells and supporting cells^{1,2}. It is not clear whether the lack of mammalian hair cell regeneration is due to an intrinsic inability of supporting cells to divide and differentiate or to an absence or blockade of regenerative signals. Here we show that post-mitotic supporting cells³ purified from the postnatal mouse cochlea retain the ability to divide and trans-differentiate into new hair cells in culture. Furthermore, we show that age-dependent changes in supporting cell proliferative capacity are due in part to changes in the ability to downregulate the cyclin-dependent kinase inhibitor *p27^{Kip1}* (also known as *Cdkn1b*). These results indicate that postnatal mammalian supporting cells are potential targets for therapeutic manipulation.

To test the capacity of post-mitotic supporting cells to re-enter the cell cycle and produce hair cells, we purified neonatal cochlear supporting cells by fluorescence-activated cell sorting (FACS) using transgenic mice expressing green fluorescent protein (GFP) under the control of the *p27^{Kip1}* promoter (Fig. 1a, b, and Supplementary Fig. 1). GFP expression is observed in all supporting cell types, including interphalangeal cells, pillar cells, Deiter's cells, and to a lesser extent Hensen's cells (Fig. 1a, b). This expression pattern largely mimics that of the endogenous *p27^{Kip1}* protein, with the exception of Claudius cells, which express lower levels of the transgene (Fig. 1b, c). Importantly, hair cells do not express *p27^{Kip1}* protein or the *p27-gfp* transgene (compare Fig. 1b with 1e). Cell counts show that more than 95% of the FACS-purified population (Fig. 1f, green box) are GFP⁺ (Fig. 1g, h, j). We also analysed expression of S100A1, which marks all supporting cells except Hensen's cells (Fig. 1d). 84% of freshly isolated *p27-GFP*⁺ cells express S100A1, similar to the abundance of this population *in vivo* (Fig. 1h–j). Conversely, the *p27-GFP*[−] population (Fig. 1f, black box) is depleted of S100A1⁺ cells (Fig. 1k–m). Quantitative polymerase chain reaction (PCR) analysis indicated that *p27-GFP*⁺ cells express enriched levels of the supporting cell-specific genes *Hes5* (ref. 4), *jagged1* (ref. 5), *p27^{Kip1}* (ref. 6) and *S100A1* (ref. 7), but not hair-cell-specific genes such as *Math1/Atoh1* (ref. 8) or *Pou3f4/Brn3c* (ref. 9) (Fig. 1n). These data indicate that the purified *p27-GFP*⁺ population is highly enriched for supporting cells.

To test whether postnatal supporting cells can divide and trans-differentiate into hair cells, we co-cultured purified *p27-GFP*⁺ supporting cells with periotic mesenchymal cells, which support the growth and differentiation of embryonic cochlear sensory progenitors¹⁰. Cultured *p27-GFP*⁺ supporting cells aggregate into small epithelial, E-cadherin⁺ islands within the mesenchymal feeder layer

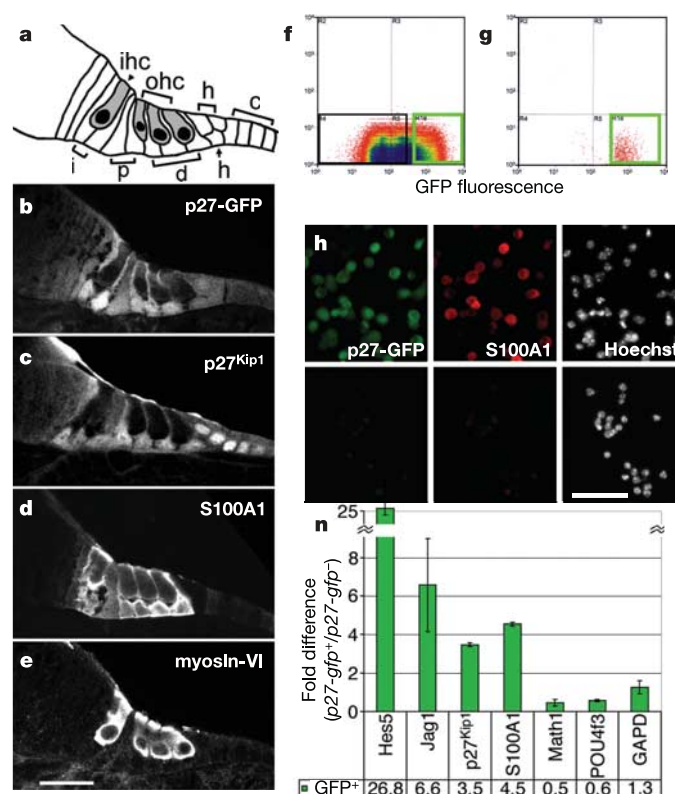


Figure 1 | Purification of neonatal cochlear supporting cells from *p27-gfp* transgenic mice. **a–e**, Organ of Corti cell types and marker expression. **a**, Schematic of sensory epithelium cell types: One inner hair cell (ihc, arrowhead) and three outer hair cells (ohc) are surrounded by different supporting cells: interphalangeal cells (i); pillar cells (p); Deiter's cells (d); Hensen's cells (h); and Claudius cells (c). **b–d**, Supporting cell-specific markers: *p27-gfp* transgene (**b**) and endogenous *p27^{Kip1}* antibody labelling (**c**); S100A1 antibody labels all supporting cells except Hensen's cells (**d**). **e**, Antibody to myosin-VI labels hair cells. **f**, FACS plot of *p27-GFP*⁺ supporting cell purification (19–22%; GFP-high, green box; GFP-low, black box). **g–m**, Verification of supporting cell purification. **g**, FACS analysis (green box; >95%). **h–j**, Microscopic analysis. >95% of Hoechst-stained nuclei (**j**) are *p27-GFP*⁺ (**h**); 84% of *p27-GFP*⁺ are S100A1⁺ (**i**). No GFP fluorescence (**k**) or S100A1 staining (**l**) was observed in the *p27-GFP*[−] population (**k–m**). **n**, Real-time quantitative PCR. Fold difference between GFP⁺ cells and GFP[−] cells. Supporting cell markers: *Hes5*, *jagged1*, *p27^{Kip1}*, *S100A1*. Hair cell markers: *Math1* and *Pou4f3*. GAPD: control. Mean $\pm$ s.e.m.; $n = 3$ triplicate samples. Scale bars: **a–e**, 20 μ m; **h–m**, 50 μ m.

¹Gonda Department of Cell and Molecular Biology, House Ear Institute, 2100 W. Third Street, Los Angeles, California 90057, USA. ²Department of Cell and Neurobiology, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA.

*These authors contributed equally to this work.

(Fig. 2a, b). Although postnatal supporting cells are post-mitotic *in vivo*³, $60 \pm 10\%$ of purified p27-GFP⁺ supporting cells down-regulate p27^{Kip1} protein after 2 days *in vitro* (DIV) (Fig. 2e), and $38 \pm 3.6\%$ of supporting cells also incorporate BrdU (Fig. 2d–f). In addition, the average number of epithelial cells per island and the fraction of supporting cells incorporating BrdU increased significantly over the first three days ($P < 0.01$, analysis of variance, ANOVA, Fig. 2g, $n = 4$ cultures, 1 and 3 DIV; $n = 8$ cultures, 2 DIV). Thus, a substantial proportion of normally post-mitotic supporting cells are able to proliferate *in vitro*, and this proliferation correlates with the downregulation of the cell cycle inhibitor p27^{Kip1}.

To test whether dividing postnatal supporting cells can trans-differentiate into hair cells, as has been observed in birds¹¹, we assayed for the hair-cell marker myosin-VI¹² in p27-GFP⁺ supporting cell cultures incubated continuously with BrdU. In a typical culture in which 870 supporting cells survived at 1 DIV, 25 ± 10 new myosin-VI⁺ cells were observed at 6 DIV ($n = 4$ cultures). 20% of these myosin-VI⁺ cells also contained BrdU (Fig. 2h–j). We observed no increase in myosin-VI⁺ cells after 6 DIV (data not shown). The presence of BrdU⁺ and BrdU[−] cells expressing myosin-VI indicates that purified postnatal mammalian supporting cells can divide and trans-differentiate into hair cells through a combination of mitotic and non-mitotic mechanisms.

To determine whether a specific sub-population of supporting

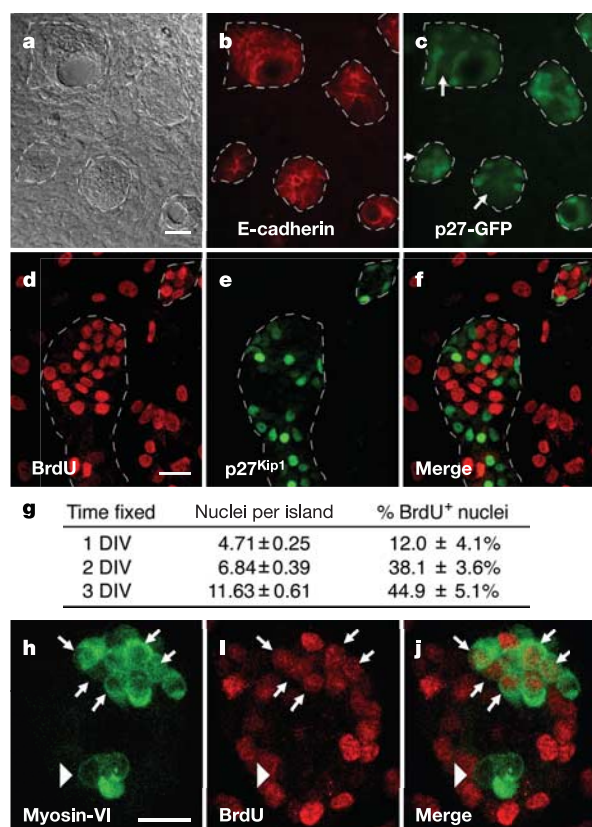


Figure 2 | Cell-cycle re-entry of supporting cells correlates with p27^{Kip1} downregulation. **a–f**, Purified P1 p27-GFP⁺ supporting cells cultured for 2 DIV. **a–c**, Epithelial islands (outlined) in a mesenchymal feeder layer (**a**), stained for the supporting cell marker E-cadherin (**b**). The p27-gfp transgene is downregulated in a subset of epithelial cells (**c**, arrows). **d–f**, BrdU staining of proliferating cells (1–2 DIV) in epithelial islands (outlined) is mutually exclusive of anti-p27^{Kip1} staining (**e**, **f**). **g**, Quantification of E-cadherin⁺ supporting cell proliferation between 1 and 3 DIV ($\pm$ s.e.m.). **h–j**, Purified supporting cells divide and trans-differentiate into BrdU⁺ hair cells by 6 DIV: anti-myosin-VI (**h**), anti-BrdU (**i**), merge (**j**); arrows indicate double-labelled cells; arrowhead indicates myosin VI-only cells. Scale bars: **a–c**, 50 μ m; **d–f**, **h–j**, 12 μ m.

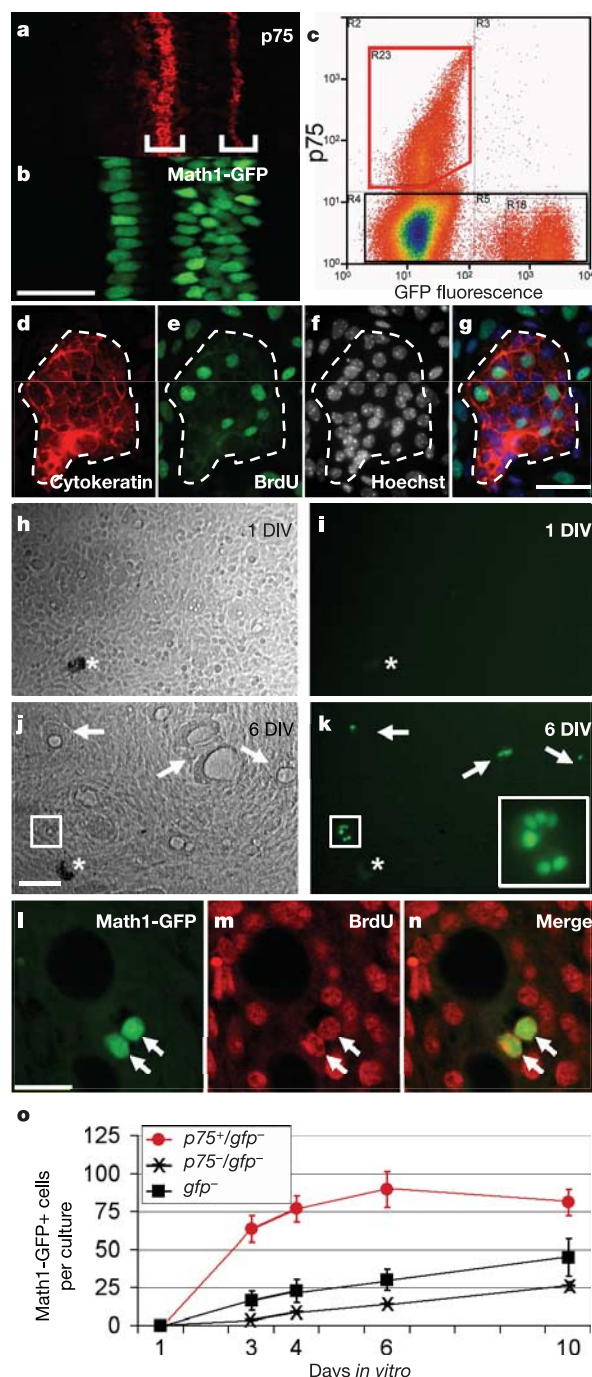


Figure 3 | Purified p75^{NGFR}⁺ pillar and/or Hensen's cells divide and generate new hair cells *in vitro*. **a, b**, Cochlea stained anti-p75^{NGFR} showing pillar cells (left bracket) and Hensen's cells (right bracket, **a**). The Math1-gfp transgene marks hair cells (**b**). **c**, FACS sorting of P2 Math1-GFP cochlear epithelial cells stained with anti-p75^{NGFR}. The y axis shows phycoerythrin-stained p75^{NGFR}. Red box: p75^{NGFR}⁺ fraction (10–12%); black box: p75^{NGFR}[−] cells. **d–g**, p75^{NGFR}⁺ supporting cells proliferate *in vitro*. A cytokeratin⁺ epithelial island (**d**, white dotted outlines) contains BrdU⁺ nuclei (**e**, 24-h pulse; **f**, Hoechst; **g**, merge). **h–k**, Differentiation of Math1-GFP⁺ hair cells in living cultures of p75^{NGFR}⁺ supporting cells. At 6 DIV but not 1 DIV, Math1-GFP⁺ are present (**h–k**, arrows). The asterisk indicates a piece of debris marking the field. **l–n**, Math1-GFP⁺ cells at 6 DIV (**l**, arrows) incorporated BrdU (**m**) during the culture period (**n**). **o**, p75^{NGFR}⁺ cells are enriched in hair-cell-generating capacity. Cochlear epithelial cells (black squares), p75^{NGFR}⁺ cells (red circles), and p75^{NGFR}[−] cells (black crosses). ($n = 4$ cultures, $\pm$ s.e.m.) Scale bars: **a, b, d–g**, 25 μ m; **h–k**, 100 μ m; **l–n**, 12 μ m.

cells is enriched for the capacity to generate hair cells, we purified pillar and Hensen's cells using antibodies to the cell surface antigen $p75^{NGFR}$, which is specifically expressed by these supporting cell subpopulations (Fig. 3a)¹³. $p75^{NGFR+}$ pillar and Hensen's cells were FACS-purified (Fig. 3c) from a transgenic mouse line in which GFP is expressed in hair cell nuclei under the control of a *Math1* enhancer^{14,15} (Fig. 3b). With this transgene, *Math1*-GFP⁺ hair cells can be eliminated during FACS purification. Moreover, it is possible to monitor the *de novo* differentiation of hair cells from the $p75^{NGFR+}$ supporting cell population with the *Math1-gfp* reporter. $p75^{NGFR+}$ supporting cells (Fig. 3c) were analysed for pillar cell markers by quantitative PCR. The purified $p75^{NGFR+}$ supporting cell fraction was highly enriched for $p75^{NGFR}$ messenger RNA and *Fgfr3*¹⁶, an independent marker expressed specifically in pillar cells (Supplementary Fig. 2a). Markers of hair cells such as *Math1* and *Pou4f3* were present at very low levels in the $p75^{NGFR+}$ population (Supplementary Fig. 2a).

We assayed $p75^{NGFR+}$ supporting cells for proliferation and hair-cell generation *in vitro*. The $p75^{NGFR+}$ supporting cell sub-fraction proliferates in culture at a rate similar to the $p27$ -GFP⁺ supporting cell population (Fig. 3d–g, and data not shown). No *Math1*-GFP⁺ cells are observed in cultures at 1 DIV (Fig. 3h, i), but at 6 DIV many new *Math1*-GFP⁺ cells are observed within the growing epithelial islands (Fig. 3j, k, arrows). 37% of these new *Math1*-GFP⁺ cells incorporated BrdU (Fig. 3l–n). By 6 DIV, an average of 54 ± 25 *Math1*-GFP⁺ cells per culture differentiated from a starting population of approximately 500 supporting cells present at 1 DIV ($n = 25$ cultures).

We wished to determine whether the $p75^{NGFR+}$ supporting cell sub-fraction was enriched for hair-cell-forming capacity relative to the other cochlear epithelial cells. We compared cultures of $p75^{NGFR+}$ supporting cells (Fig. 3o, red circles) to both $p75^{NGFR-}$ cochlear epithelial cells (Fig. 3o, crosses) and to total cochlear epithelial cells (Fig. 3o, squares), following depletion of *Math1*-GFP⁺ hair cells from each starting population. The $p75^{NGFR+}$ supporting cell fraction is significantly enriched for hair-cell-generating capacity ($P < 0.01$, Tukey's HSD, $n = 4$ cultures). Some residual capacity to generate hair cells remains in the $p75^{NGFR-}$ population (Fig. 3o, compare circles with crosses) because $p75^{NGFR-}$ cells still produce on average one-third as many *Math1*-GFP⁺ hair cells as their $p75^{NGFR+}$ counterparts after one week of culture.

During the development of the organ of Corti, hair-cell maturation occurs over a period of several weeks¹⁷. We detected markers

consistent with hair-cell maturation in cultures of $p75^{NGFR+}$ supporting cells at 14 DIV. These include myosin-VIIa¹², calretinin¹⁸, phalloidin-labelled stereocilia and espin¹⁹ (Supplementary Fig. 3). The vital dye FM1-43, which is taken up by hair cells²⁰, also specifically labelled living *Math1*-GFP⁺ cells in these cultures (Supplementary Fig. 3).

Functional and anatomical maturation of the mouse organ of Corti is not complete at birth and continues until the onset of hearing between postnatal days P10 and P14^{17,21}. To assay maturation-dependent changes in the capacity of supporting cells for cell division and trans-differentiation, we analysed the *p27-gfp* transgenic organ of Corti from P14 mice (Fig. 4). At this stage, transgene expression is confined to a subset of supporting cells including pillar, Hensen's and interphalangeal cells, but is low in Deiter's cells (Fig. 4a, b). FACS-purified cells (Fig. 4c) were 94% GFP⁺, and quantitative PCR confirmed the high level of supporting cell marker expression (Supplementary Fig. 2b). No myosin-VI⁺ cells were observed in samples of purified P14 *p27*-GFP⁺ preparations fixed at plating (not shown). Survival of P14 supporting cells was significantly lower than at P3, with an average of 100 epithelial cells surviving per culture after 48 h. In contrast to neonatal supporting cells, less than 2% of P14 supporting cells re-entered the cell cycle between 1 and 2 DIV, and *p27*^{Kip1} was not downregulated in the remaining 98% of post-mitotic supporting cells (Fig. 4d–g). Nonetheless, supporting cells still aggregated, forming small epithelial islands, and after 12 DIV an average of 2.5 myosin VI⁺ cells lacking BrdU could be observed in E-cadherin⁺ epithelial islands (Fig. 4h–j). Our cultures were free of hair cells at the time of isolation, so these cells were probably formed by the trans-differentiation of supporting cells. The yield of hair cells generated by P14 supporting cells (2.5 cells per 100 surviving plated cells) was comparable to what we observed in P3 cultures (25 cells per 870 surviving plated cells).

p27^{Kip1} is thought to play a part in maintaining the post-mitotic state of supporting cells^{6,11}. To test the role of *p27*^{Kip1} in the failure of supporting cell cycle re-entry at P14, FACS-purified supporting cells from P14 *p27*^{Kip1}-null mice were cultured and assayed for cell cycle re-entry and trans-differentiation. FACS-purification was made possible by the fact that the *p27-gfp* transgene continues to recapitulate endogenous *p27*^{Kip1} expression in the *p27*^{Kip1}-null mouse (data not shown). Supporting cells (94% pure) from P14 *p27*^{Kip1}-null mice, identified through E-cadherin expression (Fig. 4k), re-entered the cell cycle with a significantly increased frequency relative

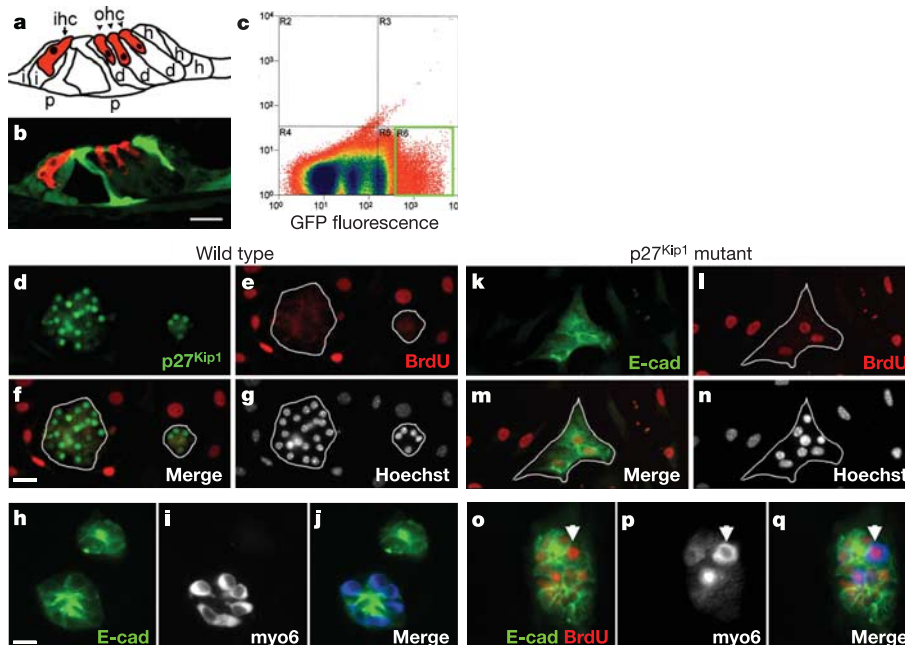


Figure 4 | P14 supporting cells generate myosin-VI⁺ cells in culture. **a**, Diagram of P14 organ of Corti, with cell types shown as in Fig. 1a. **b**, P14 *p27*-GFP transgenic cochlea: *p27*-GFP (green), myosin-VI (red). Interphalangeal cells (i), pillar cells (p) and Hensen's cells (h) express *p27*-GFP. **c**, FACS sorting of P14 *p27*-GFP supporting cells. **d–g**, 2 DIV wild-type P14 *p27*-GFP⁺ cells do not incorporate BrdU: *p27*^{Kip1} (d, f, green), BrdU (e, f, red) and Hoechst (g, white). **h–j**, 12 DIV wild-type P14 supporting cells generate small numbers of hair cells *in vitro* without dividing: E-cadherin (h, j, green), myosin-VI (i, white; j, blue) merge (j). **k–n**, 2 DIV *p27*^{Kip1}-/- P14 *p27*-GFP⁺ cells incorporate BrdU: E-cadherin (k, m, green), BrdU (l, m, red) and Hoechst (n, white). **o–q**, 12 DIV culture of *p27*^{Kip1}-/- P14 *p27*-GFP⁺ supporting cells incorporate BrdU and trans-differentiate into hair cells: E-cadherin (o, q, green), BrdU (o, q, red), myosin-VI (p, white; q, blue). Two new hair cells have BrdU⁺ nuclei (arrow). Scale bars: **b**, 10 μ m; **f** (for **d–g** and **k–n**), 25 μ m; **h** (for **h–j** and **o–q**), 12 μ m.

to their wild-type counterparts (Fig. 4k–n, $11.6 \pm 2.5\%$ versus $2.5 \pm 1.6\%$ at 1–2 DIV, $n = 9$ cultures, $P = 0.02$, Student's *t*-test). In addition, a small number of BrdU⁺ hair cells were observed after 12 DIV (Fig. 4o–q).

Mitotic regeneration has never been shown in the postnatal mammalian cochlea²². However, recent results indicate that the forced expression of *Math1* in cells of the inner ear of mammals can lead to their differentiation into hair cells^{23,24}. Unlike regeneration in birds^{25,26}, this differentiation appears to occur without concomitant supporting cell proliferation. Effective therapeutic strategies in humans are likely to require the generation of new supporting cells as well as new hair cells, so it is important to test the ability of mammalian supporting cells to divide and differentiate at postnatal times. Our work demonstrates that prospectively identified, post-mitotic, postnatal supporting cells are capable of proliferation and subsequent trans-differentiation into hair cells. In addition, the age-dependent change in the ability of supporting cells to downregulate *p27^{Kip1}* suggests a mechanistic explanation for the failure of supporting cell proliferation in response to the loss of hair cells, both in culture and *in vivo*. These results indicate that *p27^{Kip1}* represents a suitable target for hair-cell regeneration by therapeutic manipulation in postnatal supporting cells.

METHODS

Detailed methods, along with a diagram illustrating the experimental strategy used in this paper can be found in the Supplementary Methods.

Marking supporting cells by transgenesis. Supporting cells were marked by expression of GFP under the control of the *p27^{Kip1}* gene inserted into a bacterial artificial chromosome²⁷ (BAC) containing 130 kb of DNA from the *p27^{Kip1}* locus²⁸. The BAC was used to produce transgenic mice in which five copies of the transgene were present.

Purification of cochlear supporting cells by FACS. Cochlea were removed and dissociated for FACS purification. Purity of the resulting supporting cells was assessed by re-sorting of live cells, by immunohistochemistry, and by quantitative PCR.

Supporting cell culture. Purified postnatal supporting cells from wild-type or *p27^{Kip1}* mutant mice²⁹ were grown essentially as previously described¹⁰. Purified supporting cells were plated with embryonic periotic mesenchymal cells as a feeder layer, and grown in 5% CO₂/5% O₂ in DMEM/F12 (Gibco) + 2% B27 supplement (Gibco) + 1% penicillin-streptomycin (Gibco) + 5 ng ml⁻¹ EGF (Sigma) + 2.5 ng ml⁻¹ FGF2 (NIH). To label dividing supporting cells, 5'-bromo-2'-deoxyuridine (BrdU, Sigma) was included at a final concentration of 1 μM.

Quantitative PCR. Total RNA was isolated from FACS-purified samples consisting of 10,000 cells each. PCR reactions were carried out in triplicate. Relative quantification of gene expression was analysed using the $\Delta\Delta CT$ method³⁰.

Received 1 February; accepted 26 April 2006.

- Corwin, J. T. & Cotanche, D. A. Regeneration of sensory hair cells after acoustic trauma. *Science* **240**, 1772–1774 (1988).
- Ryals, B. M. & Rubel, E. W. Hair cell regeneration after acoustic trauma in adult Coturnix quail. *Science* **240**, 1774–1776 (1988).
- Ruben, R. J. Development of the inner ear of the mouse: a radioautographic study of terminal mitoses. *Acta Otolaryngol. Suppl.* **220**, 1–44 (1967).
- Zine, A. *et al.* Hes1 and Hes5 activities are required for the normal development of the hair cells in the mammalian inner ear. *J. Neurosci.* **21**, 4712–4720 (2001).
- Lewis, A. K., Frantz, G. D., Carpenter, D. A., de Sauvage, F. & Gao, W. Q. Distinct expression patterns of notch family receptors and ligands during development of the mammalian inner ear. *Mech. Dev.* **78**, 159–163 (1998).
- Chen, P. & Segil, N. *p27(Kip1)* links cell proliferation to morphogenesis in the developing organ of Corti. *Development* **126**, 1581–1590 (1999).
- Coppens, A. G., Kiss, R., Heizmann, C. W., Schafer, B. W. & Poncelet, L. Immunolocalization of the calcium binding S100A1, S100A5 and S100A6 proteins in the dog cochlea during postnatal development. *Brain Res. Dev. Brain Res.* **126**, 191–199 (2001).
- Bermingham, N. A. *et al.* Math1: an essential gene for the generation of inner ear hair cells. *Science* **284**, 1837–1841 (1999).
- Erkman, L. *et al.* Role of transcription factors Brn-3.1 and Brn-3.2 in auditory and visual system development. *Nature* **381**, 603–606 (1996).
- Doetzlhofer, A., White, P. M., Johnson, J. E., Segil, N. & Groves, A. K. *In vitro*

growth and differentiation of mammalian sensory hair cell progenitors: a requirement for EGF and periotic mesenchyme. *Dev. Biol.* **272**, 432–447 (2004).

- Roberson, D. W., Kreig, C. S. & Rubel, E. W. Light microscopic evidence that direct transdifferentiation gives rise to new hair cells in regenerating avian auditory epithelium. *Audit. Neurosci.* **2**, 195–205 (1996).
- Hasson, T. *et al.* Unconventional myosins in inner-ear sensory epithelia. *J. Cell Biol.* **1997**, 1287–1307 (1997).
- Gestwa, G. *et al.* Differential expression of trkB.T1 and trkB.T2, truncated trkC, and p75(NGFR) in the cochlea prior to hearing function. *J. Comp. Neurol.* **414**, 33–49 (1999).
- Helms, A. W., Abney, A. L., Ben-Arie, N., Zoghbi, H. Y. & Johnson, J. E. Autoregulation and multiple enhancers control Math1 expression in the developing nervous system. *Development* **127**, 1185–1196 (2000).
- Chen, P., Johnson, J. E., Zoghbi, H. Y. & Segil, N. The role of Math1 in inner ear development: Uncoupling the establishment of the sensory primordium from hair cell fate determination. *Development* **129**, 2495–2505 (2002).
- Colvin, J. S., Bohne, B. A., Harding, G. W., McEwen, D. G. & Ornitz, D. M. Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3. *Nature Genet.* **12**, 390–397 (1996).
- Lim, D. J. & Anniko, M. Developmental morphology of the mouse inner ear. A scanning electron microscopic observation. *Acta Otolaryngol. Suppl.* **422**, 1–69 (1985).
- Dechesne, C. J., Rabejac, D. & Desmadryl, G. Development of calretinin immunoreactivity in the mouse inner ear. *J. Comp. Neurol.* **346**, 517–529 (1994).
- Zheng, L. *et al.* The deaf jerker mouse has a mutation in the gene encoding the espin actin-bundling proteins of hair cell stereocilia and lacks espins. *Cell* **102**, 377–385 (2000).
- Meyers, J. R. *et al.* Lighting up the senses: FM1–43 loading of sensory cells through nonselective ion channels. *J. Neurosci.* **23**, 4054–4065 (2003).
- Pujol, R. & Hilding, D. Anatomy and physiology of the onset of auditory function. *Acta Otolaryngol.* **76**, 1–10 (1973).
- Chardin, S. & Romand, R. Regeneration and mammalian auditory hair cells. *Science* **267**, 707–709 (1995).
- Izumikawa, M. *et al.* Auditory hair cell replacement and hearing improvement by *Atoh1* gene therapy in deaf mammals. *Nature Med.* **11**, 271–276 (2005).
- Zheng, J. L. & Gao, W. Q. Overexpression of Math1 induces robust production of extra hair cells in postnatal rat inner ears. *Nature Neurosci.* **3**, 580–586 (2000).
- Warchol, M. E. & Corwin, J. T. Regenerative proliferation in organ cultures of the avian cochlea: identification of the initial progenitors and determination of the latency of the proliferative response. *J. Neurosci.* **16**, 5466–5477 (1996).
- Stone, J. S. & Rubel, E. W. Cellular studies of auditory hair cell regeneration in birds. *Proc. Natl Acad. Sci.* **97**, 11714–11721 (2000).
- Yang, X. M., Model, P. & Heintz, N. Homologous recombination-based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nature Biotechnol.* **15**, 859–865 (1997).
- Doetzlhofer, A., White, P. M., Lee, Y. S., Groves, A. & Segil, N. Prospective identification and purification of hair cell and supporting cell progenitors from the embryonic cochlea. *Brain Res.* published online 6 April 2006 (doi:PMID:16616734).
- Fero, M. L. *et al.* A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in *p27(Kip1)*-deficient mice. *Cell* **85**, 733–744 (1996).
- Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta CT)$ method. *Methods* **25**, 402–408 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We gratefully acknowledge S. Juntilla, J. Llamas and W. Makmura for animal care and genotyping; S. Chavira at the USC Flow Cytometry Section for FACS assistance, the Transgenic Mouse Core Facility at USC for generating the *p27-GFP BAC* transgenics, A. J. Hudspeth for the anti-espin antibody, T. Hasson for anti-myosin-VI and anti-myosin-VIIa, J. Johnson for Math1-GFP mice, J. Roberts for the *p27^{Kip1}-/-* mice, and E. W. Rubel for helpful advice. This work was supported by grants from the Hair Cell Regeneration Initiative of the National Organisation for Hearing Research and from the NIH. P.M.W. was supported by an NIH training grant.

Author contributions The project was conceived by A.K.G. and N.S.; experiments were planned, performed and analysed jointly by A.D. and P.M.W. with advice from A.K.G. and N.S. The *p27-gfp* transgenic mouse was constructed by Y.S.L.; the paper was written by A.D., A.K.G., N.S. and P.M.W.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.K.G. (agroves@hei.org) and to N.S. (nsegil@hei.org).

LETTERS

Clonal analysis reveals a common progenitor for thymic cortical and medullary epithelium

Simona W. Rossi¹, William E. Jenkinson¹, Graham Anderson¹ & Eric J. Jenkinson¹

The thymus provides an essential environment for the development of T cells from haemopoietic progenitors. This environment is separated into cortical and medullary regions, each containing functionally distinct epithelial populations that are important at successive stages of T-cell development and selection^{1,2}. However, the developmental origin and lineage relationships between cortical and medullary epithelial cell types remain controversial³. Here we describe a clonal assay to investigate the developmental potential of single, individually selected, thymic epithelial progenitors (marked with enhanced yellow fluorescent protein) developing within the normal architecture of the thymus. Using this approach, we show that cortical and medullary epithelial cells share a common origin in bipotent precursors, providing definitive evidence that they have a single rather than dual germ layer origin during embryogenesis. Our findings resolve a long-standing issue in thymus development, and are important in relation to the development of cell-based strategies for thymus disorders and the possibility of restoring function of the atrophied adult thymus.

At embryonic day (E) 12, the murine thymic rudiment is a relatively simple structure, consisting of a core of undifferentiated epithelial cells surrounded by a capsule of neural crest derived mesenchyme, which can be easily isolated and disaggregated into single cell suspensions for purification and analysis⁴. Ectopic transplantation of isolated E12 thymus rudiments has shown that they have considerable potential for growth and the capacity to generate normal thymic structures with cortical and medullary organization, demonstrating that all the progenitor activity required to establish a complete thymic environment is present within the isolated rudiment at this stage⁵. To define populations of epithelial progenitors in the E12 rudiment for clonal analysis, we identified all epithelial cells in disaggregated thymic rudiments by labelling with pan-cytokeratin as a universal epithelial marker⁶ and compared this with expression of EpCAM1, a cell surface marker that is expressed at different levels on both mature cortical and mature medullary epithelial cells⁷. We found that all pan-cytokeratin-positive E12 cells express comparable levels of EpCAM1 and are not differentiated into EpCAM1-high (medullary phenotype) and EpCAM1-low (cortical phenotype) populations at this stage (Fig. 1a). Thus, at E12 of gestation, EpCAM1 can be used as a pan-epithelial marker.

We next analysed expression of MTS24, a surface marker that has been reported to define a subpopulation of epithelial progenitors in the early thymus that contains the potential to generate cortical and medullary progeny^{5,8}. In contrast to the results of a previous report analysing MTS24 and cytokeratin expression individually in early thymus suspensions⁵, our use of two-colour flow cytometric analysis (allowing direct analysis of MTS24 expression within total E12 thymic epithelial cells) showed that all cytokeratin-positive cells at this stage are MTS24-positive (Fig. 1b). Collectively, our findings suggest that the epithelial population in the E12 thymic rudiment is a phenotypically homogeneous population of immature cells with

regard to expression of EpCAM1 and MTS24, and is consistent with previous findings that most E12 epithelial cells co-express cytokeratin 5 and cytokeratin 8, a phenotype suggestive of epithelial progenitors^{9,10}. On this basis we used EpCAM1 as a marker to purify cells representative of the total epithelial population in the E12 rudiment, as a starting population from which to select progenitors for further clonal analysis.

An ideal approach to clonal analysis requires that a cell be placed within its normal microenvironment where it will receive developmental cues that allow it to reveal its full potential. This requirement excludes the use of individual cells grown in isolation or in monolayer

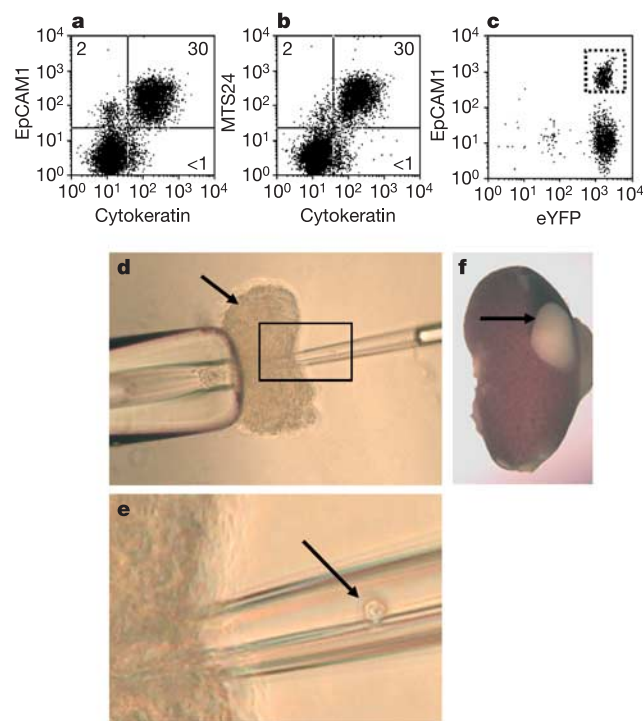


Figure 1 | A clonal assay for thymic epithelial progenitors. **a, b,** Flow cytometric analysis of E12 thymus lobes for expression of cytokeratin-positive epithelium and EpCAM1 (**a**) and MTS24 (**b**). **c,** Cell sorting strategy to isolate EpCAM1-positive, eYFP-positive cells from E12 thymus: boxed area shows the sort-gate used. **d,** Microinjection strategy: a recipient wild-type E12 thymus lobe (arrowed) is microinjected with a single EpCAM-positive eYFP-positive E12 epithelial cell; boxed area in **d** is shown at a higher magnification in **e**. Note the presence of a single cell in the microinjection pipette (arrowed in **e**). Following engraftment, microinjected lobes undergo considerable growth under the kidney capsule of adult C57BL/6 mice (**f**, arrow).

¹MRC Centre For Immune Regulation, Division of Immunity and Infection, Institute for Biomedical Research, Medical School, University of Birmingham, Birmingham B15 2TT, UK.

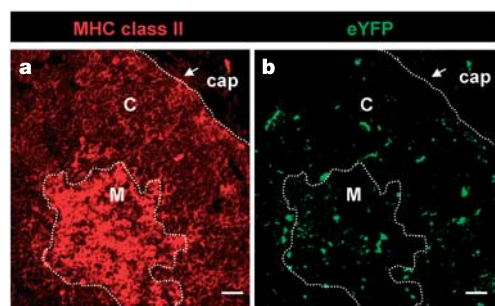


Figure 2 | A single E12 thymic epithelial cell gives rise to multiple progeny in both cortical and medullary locations *in vivo*. **a, b,** Two-colour immunofluorescence of microinjected thymus grafts 4 weeks after engraftment (**a**, MHC class II, red; **b**, eYFP, green). Multiple eYFP-positive cells derived from a single microinjected cell (green, **b**) are seen in both cortical (C) and medullary (M) areas, which are visualized by analysing MHC class II expression in the same section (**a**) to identify intensely staining medullary patches and reticular staining in cortical areas. Dotted line indicates the cortico-medullary junction, arrows indicate the capsule (cap). Scale bars, 50 µm.

cultures where, in the case of thymic epithelial cells, normal phenotype and function are perturbed¹¹. To avoid these problems, we adopted an approach in which individual epithelial precursors were re-incorporated into the normal three-dimensional architecture of an age matched developing thymus. For this purpose, EpCAM1-positive precursors were purified from isolated E12 rudiments of eYFP (enhanced yellow fluorescent protein) transgenic mice¹² to provide a starting population of intrinsically marked cells for subsequent lineage tracing (Fig. 1c). Individual cells were picked from this population using a manipulator controlled micropipette under direct visual observation and microinjected into an intact, isolated prevascular syngeneic wild-type (non-eYFP) E12 thymus rudiment (Fig. 1d and e). Microinjected lobes were then grafted under the kidney capsule of syngeneic non-eYFP mice, a site that allows for graft vascularization, growth of the grafted rudiment and the development of a normal three-dimensional cortical and medullary architecture that is capable of supporting T-cell development^{5,6,8}.

To detect progeny of individual microinjected eYFP-positive EpCAM1-positive epithelial progenitors, thymic grafts were harvested from the kidney capsule after 4 weeks, when substantial growth had taken place (Fig. 1f), and frozen sections prepared for immunohistochemistry. Cortical and medullary areas in graft sections were identified by labelling of MHC class II, which gives a delicate reticular staining pattern defining cortical regions and a more intense, confluent pattern in the medulla (Fig. 2a). Simultaneous detection of eYFP labelling in MHC class II labelled sections revealed the presence of eYFP-positive cells (which appear green) from lobes microinjected with a single eYFP progenitor. Of 13 microinjected lobes, 4 recovered grafts were found to contain eYFP-positive cells. Importantly, no eYFP signal was detected in grafts derived from control, sham injected lobes (data not shown). Interestingly, analysis of the anatomical distribution of eYFP-marked cells showed that these were present in both cortical and medullary areas (Fig. 2b), suggesting that a single progenitor can give rise to both cortical and medullary descendants.

To obtain definitive evidence on the cortical and medullary epithelial identity of eYFP-positive cells derived from a single progenitor, we next labelled sections from grafts containing the descendants of a single precursor with antibodies to a panel of markers defining cortical and medullary epithelial phenotypes. Thus, adult cortical epithelial cells are characterized by expression of cytokeratin 18 (ref. 9) and Ly51 (clone 6C3, ref. 13), both of which are absent from medullary epithelium, while medullary cells are characterized by cytokeratin-5 expression⁹ and labelling with

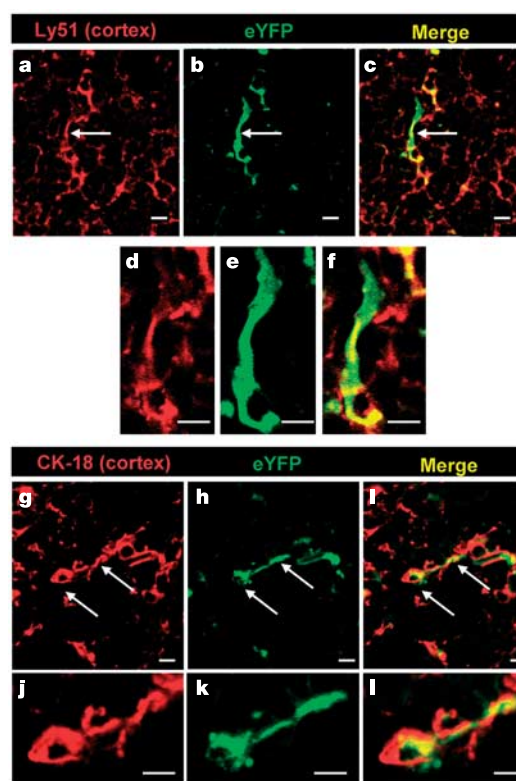


Figure 3 | Analysis of progeny of single thymic epithelial cells using cortical epithelial markers. Microinjected thymus lobes are analysed. **a–f**, Ly51 (red), eYFP (green), co-localization (yellow). **d–f**, Higher magnification of the eYFP-positive Ly51-positive cortical epithelial cell (arrowed) shown in **a–c**. **g–i**, Cytokeratin 18 (CK-18; red), eYFP (green), co-localization (yellow). **j–l**, Higher magnification cytokeratin-18-positive eYFP-positive cells (arrowed) shown in **g–i**. Scale bars, 5 µm.

antibody MTS10 (ref. 6), both of which are absent from the cortical epithelium. We therefore used Ly51 and cytokeratin 18 to identify cortical regions of grafted thymus lobes (Fig. 3). Immunohistochemical analysis shows that eYFP-positive cells are present in these areas, which also stain positive for Ly51 and cytokeratin 18 (eYFP and Ly51/cytokeratin-18 co-expression are shown as yellow, Fig. 3c, f, i and l). In sections derived from the same grafted lobe microinjected with a single eYFP progenitor, cytokeratin-5-positive and MTS10-positive cells are detectable within medullary regions, which co-label for eYFP expression (Fig. 4; yellow cells showing co-expression of eYFP and either MTS10 or cytokeratin 5 are visible in Fig. 4c, f and i). Interestingly, eYFP cells co-staining with medullary epithelial markers such as MTS10 were often present in small discrete clusters (Fig. 4d–f), reminiscent of the clonal medullary islets described in ref. 6, which range in size and were reported to contain as few as five cells. Notably, in all grafts where eYFP cells were found (4/4), both cortical and medullary derivatives of the single microinjected cell were present. From these observations we conclude that a single E12 progenitor can give rise to both cortical and medullary descendants, providing formal proof that cortical and medullary epithelial types arise from a common progenitor.

The precise developmental origin of epithelial cells in the thymus has been a long-standing controversy. Earlier studies suggested that cortical and medullary epithelium have a dual germ layer origin, being derived respectively from the ectoderm of the third pharyngeal cleft and the endoderm of the third pharyngeal pouch¹⁴. However, more recent observations have favoured the notion that thymic epithelium is exclusively derived from the endoderm of the third pouch¹⁵. Our clonal studies, showing a common origin for cortical and medullary cells, while not distinguishing between an ectodermal

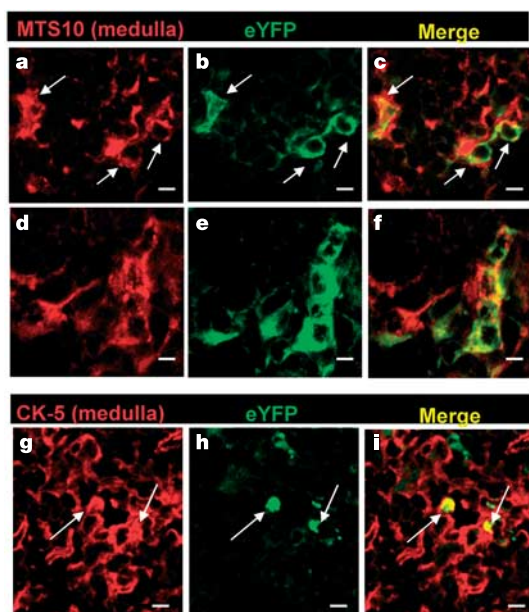


Figure 4 | Analysis of progeny of single thymic epithelial cells using medullary epithelial markers. Microinjected thymus lobes are analysed. **a–c**, **d–f**, Two examples of clusters of medullary epithelium, positive for both MTS10 (red) and eYFP (green). Co-localization is yellow. **g–i**, Expression of cytokeratin 5 (CK-5; red) and eYFP (green), with co-localization in yellow. Scale bars, 5 μ m.

or endodermal origin, provide definitive proof that these two cell types are derived from a single germ layer, and also argue against the existence of distinct cortical and medullary precursors within a single germ layer. Currently, it is not clear whether commitment to the cortical or medullary epithelial lineages is stochastic, or regulated by unknown cellular interactions. One possibility is that interactions with either surrounding neural crest mesenchyme or colonizing T-cell precursors may be a factor in determining lineage choice in bipotent thymic epithelial progenitors. However, this is unlikely in view of the fact that E12 thymic epithelial cells can undergo cortical/medullary differentiation in the absence of neural crest mesenchyme¹⁶ or a normal programme of T-cell development¹⁷. Another possibility is that lineage choice is regulated by interactions between epithelial cells, for example as a result of Notch/Notch-ligand interactions, which are known to be important in lineage choice in other epithelial systems¹⁸. As both Notch and Notch ligands are expressed on E12 and later thymic epithelial cells (ref. 4, and data not shown), we are currently investigating this possibility by inhibiting Notch signalling in E12 thymic epithelial progenitors and monitoring their subsequent development in our clonal microinjection system.

Although our observations clearly show that cortical and medullary lineages arise from common precursors in the E12 rudiment and that functionally detectable progenitors at this stage are bipotent rather than unipotent, they do not exclude the possibility that lineage restricted progenitors, possibly analogous to transit amplifying cells in other epithelial systems¹⁹, may emerge downstream of initial lineage diversification. Indeed, the clonal islets of 5–45 cells present within thymic medullary areas⁶ may arise as a result of proliferation following commitment of bipotent progenitors.

Our findings have important implications for the development of stem-cell-based strategies to restore thymus function following atrophy or age-related involution²⁰ or in the treatment of patients with defective thymus development caused by mutations in the FoxN1 gene²¹. In such strategies it will be important to achieve balanced restoration of both cortex and medulla to ensure the regeneration of T-cell immunocompetence, which is dependent on

T-cell differentiation in the cortex, without the risk of organ specific autoimmunity that is regulated by selection events in the medulla²². Our identification of bipotent progenitors for these two populations and their use at a population level to generate new thymus tissue suggests this may be feasible. The clonal approach we describe will also be useful in determining whether or not bipotent or unipotent epithelial progenitors persist in the adult among candidate populations defined by expression of the marker MTS24 (refs 5, 8) or displaying side population characteristics revealed by Hoechst 33342 staining²³. Finally, our experimental approach of injecting individual genetically marked progenitors into isolated embryonic organ anlage, followed by ectopic grafting to sites that support *in vivo* organogenesis, provides a tool that may be applicable to the clonal analysis of lineage relationships between stromal populations in a number of organ systems.

METHODS

Antibodies, histology and flow cytometry. To analyse the phenotype of epithelial cells present within the E12 thymus, freshly dissected lobes were disaggregated using trypsin as described¹⁶. For flow cytometry, we used APC conjugated anti-EpCAM-1 (gift from M. Kim, originally from A. Farr), and biotinylated MTS24 (gift from R. Boyd) followed by streptavidin-PE (Pharmingen). For analysis of cytokeratin expression, freshly disaggregated E12 thymus cell suspensions were permeabilized (PermeaFix-kit, BD Bioscience) and then labelled with FITC anti-pan-cytokeratin (C-11, Sigma). Isotype matched controls were used to set up background levels of staining. Analysis was performed using a BD LSR flow cytometer with forward and side scatter gates set so as to exclude non-viable cells. For histological analysis of grafted thymic tissue, intact grafts were incubated for 24 h in 4% paraformaldehyde (Sigma), and then processed through an increasing gradient of sucrose (12%, 15%, 18% and 30%) each time for 24 h. Tissues were embedded in OCT medium (Tissue-Tek, Sakura Finetec), snap frozen, and sections were cut at 6 μ m thickness using a cryostat. Sections were then labelled with biotinylated anti-MHC class II (IAb, clone AF6-120.1, Pharmingen), the cortical epithelial markers cytokeratin 18 (gift from B. Lane) and anti-Ly51 (clone 6C3, biotinylated, gift from M. Kim), and the medullary epithelial markers MTS10 (gift from R. Boyd) and anti-cytokeratin 5 (Covance). Second step reagents used were: anti-rat IgG AlexaFluor594, anti-rat IgM AlexaFluor594, anti-rabbit AlexaFluor594, streptavidin-AlexaFluor594 (Molecular Probes). Images were recorded using a Axiovert 100M confocal microscope (Carl Zeiss) and the software LSM510.

Microinjection of single thymic epithelial progenitors. EpCAM1-positive epithelial cells were isolated from E12 gestation eYFP C57/BL6 transgenic mice¹² at a purity greater than 99%, using a Mo-Flo high-speed cell sorter (DakoCytomations). Single eYFP-positive, EpCAM1-positive cells were hand-picked and microinjected into E12 C57/BL6 thymus lobes using a Nikon Phase Contrast microscope (ELWD0.3), Microtec holding pipette (Micro Instruments), Cell Tram Vario injector (Eppendorf) and needles self-customized using a PC-10 pipette puller (Narishige). Microinjected lobes were then grafted under the kidney capsule of adult C57BL/6 mice^{5,6,8} and recovered after 4 weeks.

Received 20 December 2005; accepted 10 April 2006.

- Petrie, H. T. Cell migration and the control of post-natal T-cell lymphopoiesis in the thymus. *Nature Rev. Immunol.* **3**, 859–866 (2003).
- Anderson, G. & Jenkinson, E. J. Lymphostromal interactions in thymic development and function. *Nature Rev. Immunol.* **1**, 31–40 (2001).
- Manley, N. R. & Blackburn, C. C. A developmental look at thymus organogenesis: where do the non-hematopoietic cells in the thymus come from? *Curr. Opin. Immunol.* **15**, 225–232 (2003).
- Harman, B. C. *et al.* T/B lineage choice occurs prior to intrathymic Notch signaling. *Blood* **106**, 886–892 (2005).
- Bennett, A. R. *et al.* Identification and characterization of thymic epithelial progenitor cells. *Immunity* **16**, 803–814 (2002).
- Rodewald, H. R., Paul, S., Haller, C., Bluethmann, H. & Blum, C. Thymus medulla consisting of epithelial islets each derived from a single progenitor. *Nature* **414**, 763–768 (2001).
- Farr, A. G., Nelson, A., Truex, J. & Hosier, S. Epithelial heterogeneity in the murine thymus: a cell surface glycoprotein expressed by subcapsular and medullary epithelium. *J. Histochem. Cytochem.* **39**, 645–653 (1991).
- Gill, J., Malin, M. A., Hollander, G. A. & Boyd, R. L. Generation of a complete thymic microenvironment by MTS24(+) thymic epithelial cells. *Nature Immunol.* **3**, 635–642 (2002).
- Klug, D. B. *et al.* Interdependence of cortical thymic epithelial cell

- differentiation and T-lineage commitment. *Proc. Natl Acad. Sci. USA* **95**, 11822–11827 (1998).
10. Klug, D. B., Carter, C., Gimenez-Conti, I. B. & Richie, E. R. Thymocyte-independent and thymocyte-dependent phases of epithelial patterning in the fetal thymus. *J. Immunol.* **169**, 2842–2845 (2002).
 11. Anderson, G., Hare, K. J., Platt, N. & Jenkinson, E. J. Discrimination between maintenance- and differentiation-inducing signals during initial and intermediate stages of positive selection. *Eur. J. Immunol.* **27**, 1838–1842 (1997).
 12. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4–12 (2001).
 13. Gray, D. H., Chidgey, A. P. & Boyd, R. L. Analysis of thymic stromal cell populations using flow cytometry. *J. Immunol. Meth.* **260**, 15–28 (2002).
 14. Cordier, A. C. & Haumont, S. M. Development of thymus, parathyroids and ultimobranchial bodies in NMRI and nude mice. *Am. J. Anat.* **157**, 227–263 (1980).
 15. Gordon, J. *et al.* Functional evidence for a single endodermal origin for the thymic epithelium. *Nature Immunol.* **5**, 546–553 (2004).
 16. Jenkinson, W. E., Jenkinson, E. J. & Anderson, G. Differential requirement for mesenchyme in the proliferation and maturation of thymic epithelial progenitors. *J. Exp. Med.* **198**, 325–332 (2003).
 17. Jenkinson, W. E., Rossi, S. W., Jenkinson, E. J. & Anderson, G. Development of functional thymic epithelial cells occurs independently of lymphostromal interactions. *Mech. Dev.* **122**, 1294–1299 (2005).
 18. Yamamoto, N., Tanigaki, K., Han, H., Hsieh, H. & Honjo, T. Notch/RBP-J signaling regulates epidermis/hair fate determination of hair follicular stem cells. *Curr. Biol.* **13**, 333–338 (2003).
 19. Jones, P. H., Harper, S. & Watt, F. M. Stem cell patterning and fate in human epidermis. *Cell* **80**, 83–93 (1995).
 20. Wils, E. J. & Cornelissen, J. J. Thymopoiesis following allogeneic stem cell transplantation: new possibilities for improvement. *Blood Rev.* **19**, 89–98 (2005).
 21. Pignata, C. *et al.* Human equivalent of the mouse nude/SCID phenotype: long-term evaluation of immunologic reconstitution after bone marrow transplantation. *Blood* **97**, 880–885 (2001).
 22. Derbinski, J., Schulte, A., Kyewski, B. & Klein, L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nature Immunol.* **2**, 1032–1039 (2001).
 23. Smalley, M., Titley, I. & Ashworth, A. An improved definition of mouse mammary epithelial side population cells. *Cytotherapy* **7**, 495–508 (2005).
- Acknowledgements** We thank R. Bird for cell sorting, and S. Parnell and T. Jones for technical assistance. This work was supported by the European Union (FP6 Eurothymaide Constortium) and the Medical Research Council, UK.
- Author Contributions** S.W.R. performed the experiments shown in Figs 2–4, S.W.R. and W.E.J. performed those shown in Fig. 1. G.A. and E.J.J. designed the experiments and wrote the paper. All authors discussed the results and commented on the manuscript.
- Author Information** Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.A. (g.anderson@bham.ac.uk).

LETTERS

Formation of a functional thymus initiated by a postnatal epithelial progenitor cell

Conrad C. Bleul¹, Tatiana Corbeaux¹, Alexander Reuter^{1†}, Paul Fisch², Jürgen Schulte Mönting³ & Thomas Boehm¹

The thymus is essential for the generation of self-tolerant effector and regulatory T cells. Intrathymic T-cell development requires an intact stromal microenvironment, of which thymic epithelial cells (TECs) constitute a major part^{1–3}. For instance, cell-autonomous genetic defects of forkhead box N1 (*Foxn1*)⁴ and autoimmune regulator (*Aire*)⁵ in thymic epithelial cells cause primary immunodeficiency and autoimmunity, respectively. During development, the thymic epithelial rudiment gives rise to two major compartments, the cortex and medulla. Cortical TECs positively select T cells⁶, whereas medullary TECs are involved in negative selection of potentially autoreactive T cells⁷. It has long been unclear whether these two morphologically and functionally distinct types of epithelial cells arise from a common bi-potent progenitor cell⁸ and whether such progenitors are still present in the postnatal period. Here, using *in vivo* cell lineage analysis in mice, we demonstrate the presence of a common progenitor of cortical and medullary TECs after birth. To probe the function of postnatal progenitors, a conditional mutant allele of *Foxn1* was reverted to wild-type function in single epithelial cells *in vivo*. This led to the formation of small thymic lobules containing both cortical and medullary areas that supported normal thymopoiesis. Thus, single epithelial progenitor cells can give rise to a complete and functional thymic microenvironment, suggesting that cell-based therapies could be developed for thymus disorders.

The thymus is the primary lymphoid organ that generates a diverse and self-compatible T-cell repertoire. The epithelial component of the thymic stroma is essential for intrathymic T-cell development^{2,3}. In mice homozygous for loss-of-function mutations in the gene encoding the transcription factor *Foxn1*⁹, TECs fail to differentiate^{4,10} and thymopoiesis is completely blocked, causing severe immunodeficiency¹¹. In the normal thymus, TECs found in the cortex and the medulla are functionally distinct. For example, medullary epithelial cells are required for negative selection, failure of which leads to autoimmunity^{5,12,13}. Therefore, TECs might be an attractive source of cell-based strategies to correct or modulate autoimmune disorders⁸.

However, much remains to be learned about their development and function. For instance, an epithelial cell population isolated from an embryonic thymus gives rise to a functional microenvironment consisting of both medullary and cortical compartments when transplanted into ectopic sites^{14,15}, but it is unclear whether this activity resides in one cell type or is the result of the joint activity of many different cell types. Furthermore, it is unclear whether bi-potent progenitors exist in the postnatal thymus. Here, we examine two questions. First, we ask whether cortical and medullary epithelial cell types are derived from one common progenitor and whether such epithelial progenitor cells exist in postnatal thymus tissue. Second, we ask whether single such progenitor cells can give rise

to functionally competent thymic tissue supporting normal thymopoiesis.

To answer the first question, we developed a lineage-tracing procedure based on the *Cre/loxP* system to follow the fate of genetically marked TECs. As a read-out for Cre activity we used the well-established *Rosa26R-eYFP* (where eYFP is enhanced yellow fluorescent protein) reporter¹⁶ system that supplies a fluorescent protein only after Cre-mediated chromosomal rearrangement at a ubiquitously expressed genomic locus. Cre activity was provided by a modified Cre-recombinase, driven by the human *K14* promoter (*hK14::Cre-ERT2*) (where ER is human oestrogen receptor)¹⁷. Two features made the latter component particularly attractive for our purpose. First, the *K14* promoter is active in epithelial progenitor cells¹⁸ and so we expected that it might be active also in thymic epithelial precursor cells. Second, Cre-ERT2 exhibits a low level of activity even in the absence of tamoxifen induction (data not shown), effectively creating somatic mosaics in epithelial compartments.

Although no yellow cells were found in the thymic lobes of newborn (postpartum day zero, P0) *hK14::Cre-ERT2;Rosa26R-eYFP* mice ($n = 10$), about half the mice (15/29) contained yellow cells in the thymic epithelial compartment at P14 (Supplementary Table S1). As mice grew older, the proportion of mice with yellow TECs steadily increased, as did the number of yellow cells per thymus. However, at P14, the number of marked cells per thymus was low, ranging from 3 to 146 cells per total thymus and thus comprised only a minor fraction ($\ll 0.1\%$) of TECs (Supplementary Table S1). In thymus tissue, yellow cells were found to be clustered rather than evenly distributed. Indeed, the observed distribution of yellow cells per unit volume (that is, per section), each comprising about 1–2% of thymic tissue, is incompatible with a Poisson distribution expected for random distribution (Supplementary Table S1, $P < 0.001$, Sign test). This indicated that yellow cells did not arise independently and that cell clusters were the result of single recombination events occurring in cells with proliferative potential.

Of the 58 thymic lobes derived from the 29 animals that we completely sectioned, 21 circumscribed areas (clusters) contained yellow cells; nine single cells were also found (Supplementary Fig. S1a). Four cell clusters occurred in the medulla, each without evidence of associated cortical cells (Fig. 1a; Supplementary Fig. S1); these clusters are reminiscent of the medullary islets described earlier¹⁹ and their observation provides independent evidence for a medullary epithelial progenitor cell. In ten instances, yellow cells occurred only in the cortex but not in the medulla, and nine of those ten instances consisted of a single cell only (Fig. 1b; Supplementary Fig. S1). One cortical cluster was observed (Fig. 1c), consisting of cells arranged concentrically around a medullary islet. This finding is compatible with the existence of a dedicated cortical progenitor cell.

¹Department of Developmental Immunology, Max-Planck Institute of Immunobiology, Stuebeweg 51, D-79108 Freiburg, Germany. ²Department of Pathology, University of Freiburg, Breisacher Strasse 115a, D-79110 Freiburg, Germany. ³Institute of Medical Biometry and Informatics, University of Freiburg, Stefan-Meier-Strasse 26, D-79104 Freiburg, Germany. [†]Present address: Department of Biology, University of Konstanz, D-78457 Konstanz, Germany.

Sixteen of 21 cell clusters (76%) contained cells of both cortical and medullary phenotypes spanning the cortico-medullary junction (Fig. 1d). These patterns suggest that TEC progenitors exhibiting yellow fluorescence after Cre-mediated recombination at the *Rosa26* locus give rise to groups of genetically related progeny that are found in close proximity to each other. Indeed, mathematical modelling reproduces the appearance of clustered progeny from single progenitors and the results indicate that biologically meaningful parameters for proliferation probabilities of progenitor and differentiated cells and generation times can explain the observed patterns (Supplementary Information and Table S2). Interestingly, the distribution of labelled progeny derived from eYFP-positive/MHC class-II-positive embryonic day (E)12.5 embryonic TECs added to reaggregates of thymic tissue and transplanted into ectopic sites is strikingly similar to the pattern observed in the postnatal lineage tracing experiment (Supplementary Fig. S2). This is compatible with the notion that the differentiation of embryonic and postnatal TECs follows similar rules.

Collectively, the above analyses strongly suggested the presence of bi- and uni-potent progenitor cells for the cortical and medullary compartments, respectively. However, owing to the experimental design of creating somatic mosaics, it was impossible to obtain proof for the functional competence of such progenitors beyond their morphological and immunohistochemical characteristics. We therefore devised a genetic test to examine the proposed differentiation model in functional terms. The design of this experiment was based on the assumption that *Foxn1*-deficient TECs are maintained in an undifferentiated state in the alymphoid rudiment^{4,10}.

We sought to create a null allele of *Foxn1* that could be reverted to the wild-type configuration upon the action of Cre recombinase. Owing to the unique properties of the *hK14::Cre-ERT2* transgene described above, we expected to be able to examine the functional consequences arising from the reversion of *Foxn1* in single epithelial cells within a *Foxn1*-deficient epithelial rudiment. We first confirmed that the *hK14::Cre-ERT2* transgene was also active in *Foxn1*^{-/-}

epithelial cells using mice transgenic for *hK14::Cre-ERT2* and *Rosa26R-lacZ*²⁰ on the *Foxn1*-deficient *nu/nu* background (Fig. 2a). In the thymic rudiment of *hK14::Cre-ERT2;Rosa26R-lacZ;nu/nu* mice, expression of β -galactosidase was confined to rare, often single, cells in an entire rudiment, indicating that Cre-mediated recombination occurred as a singular event. Thus, for subsequent experiments using the *hK14::Cre-ERT2* transgene, differentiated progeny of reverted epithelial cells could be expected to originate from single TEC.

The revertible allele of *Foxn1* was created by insertion of a cassette containing splice acceptor and polyadenylation sites into the intron between exons 6 and 7 of the mouse *Foxn1* gene (Fig. 2b). This insertion was modelled according to a rat *nude* allele²¹ to corrupt the normal splicing process. The formation of truncated non-functional messenger RNAs ultimately resulted in a *nude* phenotype in mice homozygous for this allele (Supplementary Fig. S3a). Excision of the cassette by Cre recombinase was expected to restore normal splicing; indeed, a wild-type phenotype was observed after deletion of the cassette in the germ-line (Supplementary Fig. S3b, c). When mice homozygous for the *Foxn1*^{SA2} allele were made transgenic for the *hK14::Cre-ERT2* construct, peripheral T cells could be detected (Fig. 2c) in a progressively larger proportion of the mice with increasing age. For instance, in one cohort, at 4 weeks of age, 8/18 mice had peripheral T cells, and another five became positive within the following month. The presence of peripheral T cells above the background of T cells observed in *nude* mice was absolutely dependent on the presence of the *hK14::Cre-ERT2* transgene, indicating that the *Foxn1*^{SA2} allele itself is not leaky.

Peripheral T cells in the reverter mice, although fewer in number than in wild-type mice, exhibited high surface levels of T-cell receptor (TCR) β (Fig. 2c), showed a polyclonal T-cell repertoire (Supplementary Fig. S4), upregulated activation markers upon CD3 ϵ cross-linking (Supplementary Fig. S5) and re-established the immune response to a T-cell-dependent antigen (Supplementary Fig. S6), as well as reactivity in mixed lymphocyte reactions (data not

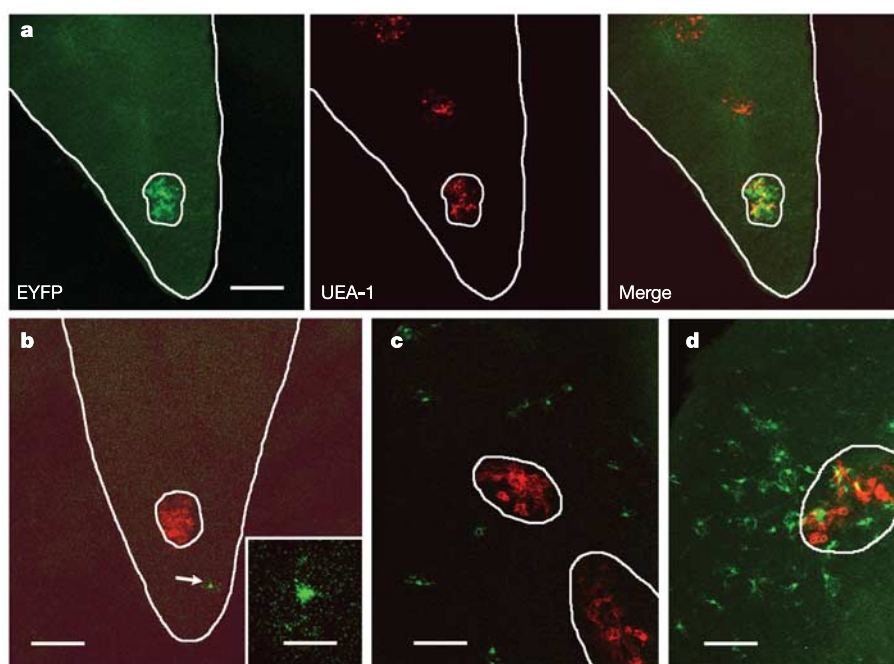


Figure 1 | Examples of cell clusters observed in lineage tracing analysis. **a**, Medullary islet observed via eYFP fluorescence (left); same section stained with UEA-1, which marks a subset of mature medullary cells (middle); merged images of eYFP and UEA-1 (right). Outline of section and genetically marked medullary islet indicated in white. **b**, Single marked

cortical cell (merged images). **c**, Halo of cortical epithelial cells adjacent to medullary islets (red) (merged images). Cortical clusters are more dispersed than medullary islets. **d**, Group of cells spanning the cortico-medullary junction (merged images). Scale bars, 100 μ m (**a**), 100 μ m (**b**), 20 μ m (**b** inset), 50 μ m (**c**), 50 μ m (**d**).

shown). When the thymic rudiment was analysed by histology in such mice, small areas of thymic tissue were apparent adjacent to alymphoid cysts (Fig. 3a). These neo-thymi contained well-defined central medullary areas enclosed by cortical spheres (Fig. 3b) and

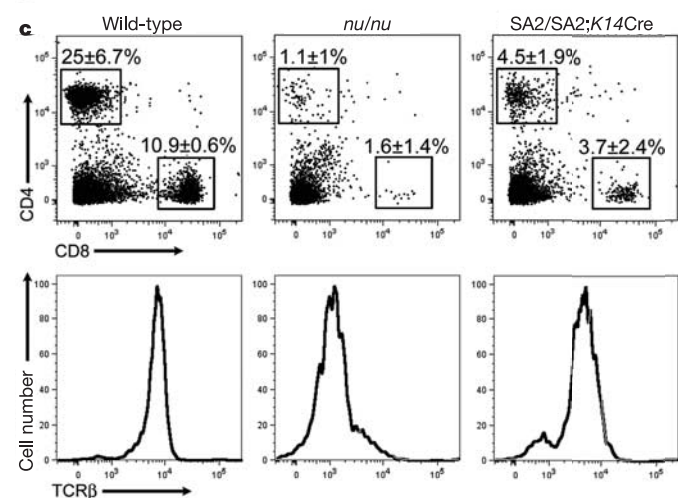
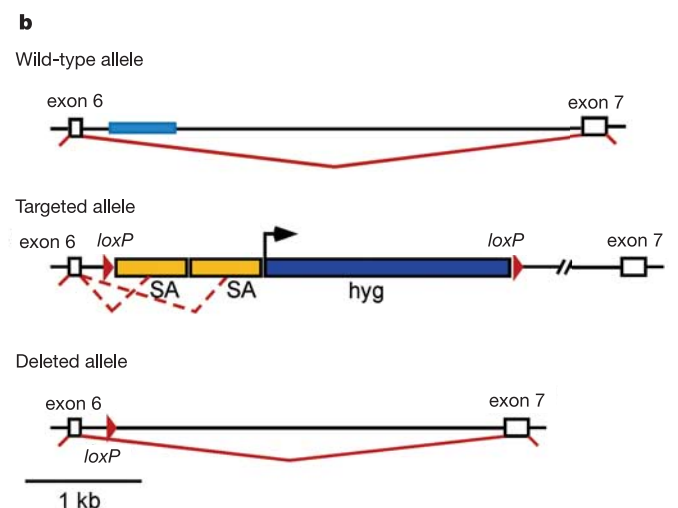
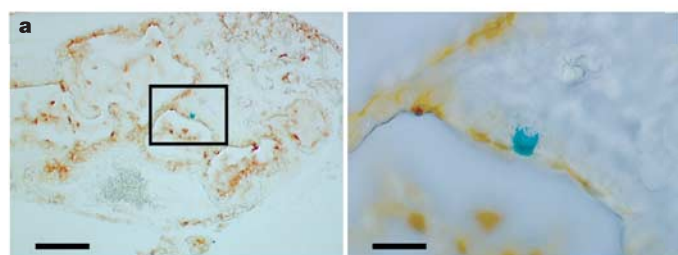


Figure 2 | A revertable *Foxn1* allele (*Foxn1*^{SA2}). **a**, Histological section of thymic rudiment in *hK14::Cre-ERT2*;*Rosa26R-lacZ*;*nu/nu* mice stained with an anti-cytokeratin K8 antibody (brownish stain) and developed for β-galactosidase expression (blue stain). Right panel, close-up of the indicated region in the left panel; note the single positive cell in the wall of the cyst. Scale bars, 100 μm, left panel; 25 μm, right panel. **b**, Structure of the *Foxn1* locus before (upper line) and after knock-in (middle line) of the SA2 cassette (yellow) and the hygromycin selectable marker (dark blue). Cre-mediated excision of this cassette (bottom line) deletes a small region (indicated by light blue rectangle) in the intron. Red arrowheads, *loxP* sites; splicing pattern indicated by red lines. **c**, Top panels, analysis of splenocytes for CD4-positive and CD8-positive T cells in 8–12-week-old wild-type, nude (*nu/nu*), and *Foxn1*^{SA2}/*Foxn1*^{SA2};*hK14::Cre-ERT2* mice (*SA2/SA2;K14Cre*) (average ± standard deviation). Bottom panels, surface levels of TCRβ for CD4-positive cells.

supported normal T-cell differentiation (Fig. 3c, d; Supplementary Fig. S7a) for at least 21 months (the latest time point examined; Supplementary Fig. S7b). The epithelium in neo-thymi expresses the autoimmune regulator *Aire* and peripheral self antigens (Supplementary Fig. 8), suggesting that central tolerance mechanisms are in place.

Interestingly, no thymic mass consisting of either cortical or medullary epithelium was observed, indicating that to support productive thymopoiesis, the reversion of *Foxn1* function in single cells occurs in a cell type that is positioned upstream of cortical and medullary precursors in the differentiation hierarchy. This experiment shows how epithelial stroma emanating from single progenitor cells gives rise to a microenvironment that supports the immigration, differentiation and emigration of haematopoietic cells. As expected for a cumulative random reversion process, the number and size of

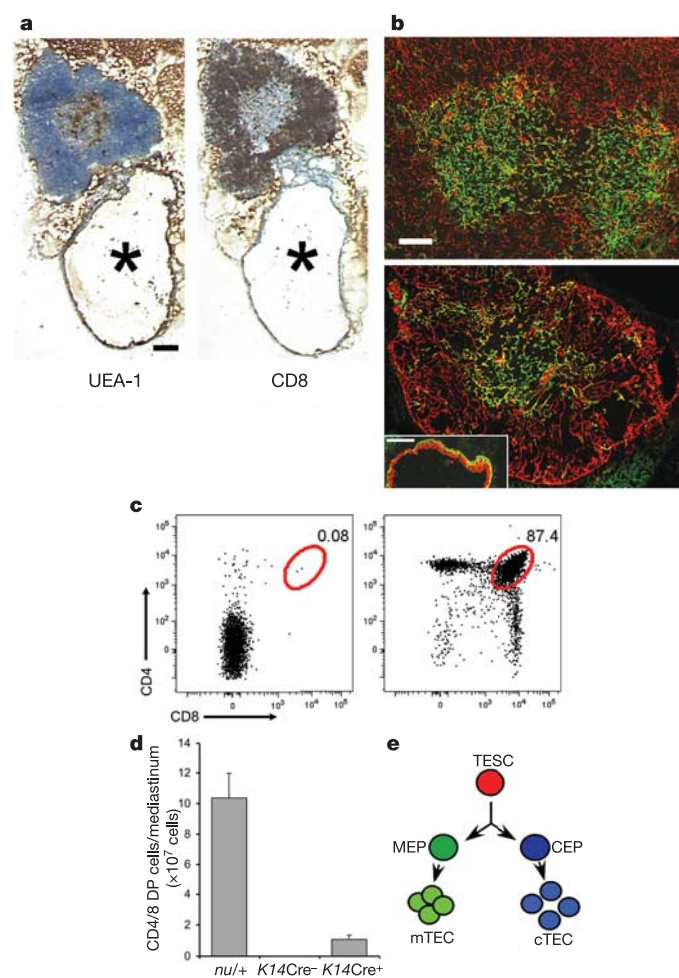


Figure 3 | Postnatal thymopoiesis in *Foxn1*^{SA2}/*Foxn1*^{SA2};*hK14::Cre-ERT2* mice. **a**, Histological section of neo-thymus stained with UEA-1 (left) and anti-CD8 (right). **b**, Anti-cytokeratin K5 (green, specific for medullary TECs) and anti-cytokeratin K8 (red, specific for cortical TECs) staining; cystic rudiment (inset). K5/K8-double positive cells appear yellow. Scale bars, 100 μm. **c**, Analysis of lymphocytes from the mediastinum of a *Foxn1*^{SA2}/*Foxn1*^{SA2} seven-week-old control mouse (left) and a *Foxn1*^{SA2}/*Foxn1*^{SA2};*hK14::Cre-ERT2* littermate (right). The double-negative cells in the left panel are B cells. **d**, Enumeration of CD4/CD8-double-positive cells in the mediastinum of 7–10-week-old *Foxn1*^{SA2}/*Foxn1*^{SA2} mice with or without the *hK14::Cre-ERT2* (*K14Cre*) transgene; control, *nu/+* mice (average ± standard deviation). **e**, Proposed scheme of TEC differentiation. A bi-potent progenitor of cortical and medullary epithelium (TESC, thymic epithelial stem cell) gives rise to distinct intermediary progenitor cells (MEP, medullary epithelial progenitor; CEP, cortical epithelial progenitor) and finally to mature medullary and cortical TECs (mTECs and cTECs).

thymic lobules observed in thymic rudiments increased in older mice (Supplementary Fig. S9a); furthermore, the age of analysed mice correlated with increasing numbers of TECs, CD4/CD8 double-positive thymocytes and peripheral T-cell numbers (Supplementary Fig. 9b, c). We also observed re-growth of hair in progressively larger patches of skin in reverter mice, indicating restoration of hair keratinization as a consequence of normalized *Foxn1* function.

Our results provide answers to a number of key problems in TEC biology. We show that the embryonic and adult TEC compartments contain bi-potent progenitor cells that give rise to TECs with both cortical and medullary phenotypes. Our results confirm the presence of medullary precursors¹⁹ and suggest the presence of an equivalent progenitor for the cortical compartment. We cannot determine the relative proportion of these types of progenitor cells in the postnatal thymus, because the method we use here for their detection depends on the (unknown) relative activity of the transgenic construct supplying Cre activity in the different precursor types. Nevertheless, our findings allow us to propose a differentiation scheme of TECs (Fig. 3e): lineage tracing and genetic reversion analysis both suggest that a *Foxn1*-dependent bi-potent precursor initially gives rise to cortical and medullary precursors which then form small clusters of differentiated progeny. Interestingly, the reversion experiment also indicates that direct interaction of ectoderm with endoderm is not required for full differentiation of adult TECs, because mediastinal thymic rudiments are no longer in contact with the cells of this germ layer. Our results also indicate that the bi-potent TEC progenitor cell expresses *Foxn1*.

A surprising finding was that despite the absence of *Foxn1*, epithelial precursors in thymic rudiments survive into the adult period, because they can be induced postnatally to differentiate normally by re-supply of *Foxn1* function. This finding provides functional proof for earlier suggestions of the presence of an immature epithelium in *Foxn1*-deficient thymic rudiments¹⁰. It also shows how a functional epithelial compartment arising from single progenitor cells is capable of organizing fully competent thymopoietic niches. The supply of *Foxn1* is necessary and sufficient to direct the differentiation of immature epithelial cells into fully mature and functionally competent TECs, so the delivery of *Foxn1* to even single progenitor cells of the cystic epithelial rudiment of human *FOXN1*^{-/-} patients^{22,23} (for instance, through viral vectors) may result in clinically relevant and long-lasting thymopoiesis that bone marrow transplantation failed to achieve²⁴. Our results thus suggest a treatment method for this particular kind of primary immunodeficiency.

The presence of a bi-potent progenitor cell in the embryonic thymic epithelial compartment has been unequivocally demonstrated²⁵; our results indicate that such a progenitor also exists in the postnatal thymus. Further work should focus on the isolation of this rare cell type from adult tissues with the possibility of developing cell-based therapies to modify the outcome of autoimmune disorders in adults.

METHODS

More detailed methods are described in the Supplementary Information.

Mice. The Cre-transgenic mouse strain for the ubiquitous deletion of *loxP*-flanked gene segments (P_{BI-2}::Cre; 'Cre-deleter' (ref. 26)) was the kind gift of K. Rajewsky. The Cre reporter strains produced by targeted insertion of *lacZ*²⁰ or eYFP¹⁶ into the *Rosa26* locus were kindly provided by P. Soriano and S. Srinivas, respectively. Mice expressing eYFP in all cells were generated by crossing *Rosa26*-eYFP mice with the Cre-deleter strain. The transgenic mouse line driving a modified Cre-protein (Cre-ERT2) under the control of the *hK14* promoter/enhancer¹⁷ was the kind gift of P. Chambon.

The *Foxn1*^{SA2} knock-in allele was created as follows. The targeting construct was assembled in a step-wise fashion. Into the *Xba*I site of the plox2d plasmid²⁷ that is positioned between two tandemly arranged *loxP* sites we inserted two tandemly arranged copies of a 605-base pair (bp) fragment (one copy of this fragment corresponds to nucleotides (nt) 321 to 925 in GenBank accession number Y08910; ref. 21.) each containing splice acceptor and polyadenylation signal sequences to give plasmid pP4SA3. This fragment is derived from a

transposon insertion into the rat *Foxn1* gene identified in the rat *nu*^{tmuN} strain²¹. A triplicated version of this cassette was inserted into plasmid pKSB22A (a modified plasmid derivative of the λ GET exon-trapping vector²⁸) to give plasmid p22SA3i and we confirmed its desired effect on splicing and polyadenylation after transfection into COS7 cells (data not shown). This indicates that the cassette may be generally useful for the construction of revertable alleles.

Immediately upstream of the downstream *loxP* site of pP4SA3, a 2.1-kilobase (kb) *Bgl*III fragment derived from pHA58 (the kind gift of A. Berns; ref. 29.) encoding a hygromycin resistance cassette was cloned. Two fragments with homologies to the mouse *Foxn1* locus were positioned up- and downstream, respectively, of the two *loxP* sites in the final targeting vector. The upstream short homology arm corresponds to nt 36917 to 39877, and the downstream long homology arm corresponds to nt 40433 to 47098 in GenBank accession number Y12488 (ref. 30). The final targeting construct (pP4SA2Whn) contains a deletion of 555 bp of intronic sequence between exon 6 (occurring at nt 39555 to 39651 in GenBank accession number Y12488) and exon 7 (occurring at nt 43958 to 33156 in GenBank accession number Y12488) of the *Foxn1* gene to facilitate the detection of Cre-mediated removal of the SA2-hyg cassette (Fig. 2b). This vector was linearized with *Xho*I and electroporated into R1 embryonic stem cells. Hyg-resistant cells were screened for homologous recombination by Southern blotting using *Bam*HI-digested DNA and a probe derived from exon 3 of *Foxn1* (nt 34186 to 34467); the wild-type allele corresponds to 18.2 kb, the targeted allele to about 11 kb. Several correctly targeted clones were injected into blastocysts, and the chimaeras backcrossed to C57BL/6 mice for germline transmission.

Heterozygous mice were intercrossed to yield homozygous mutant mice and their phenotype analysed as described⁴; as expected, the knock-in allele was allelic to the *nu* allele (data not shown) and indistinguishable in phenotype (Supplementary Fig. S3). It was concluded that the *Foxn1*-SA2hyg knock-in allele (henceforth designated *Foxn1*^{SA2}) represents a null allele of *Foxn1*. To confirm the functionality of the *loxP* sites, these mice were crossed to the Cre-deleter strain; the phenotype of mice homozygous for the deleted allele was normal (Supplementary Fig. S3).

Characterization of TECs in lineage tracing experiments. Fluorescence signals emanating from eYFP-expressing cells and from immunohistochemical stains (UEA-1; K8; MTS-10) were recorded in tissue slices (50 μ m) using a Leica TCS SP2 UV confocal microscope system. Background fluorescence was distinguished from eYFP fluorescence using Leica Spectral Dye Separation software. The average diameter of clusters composed of yellow cells was found to be $160 \pm 115 \mu$ m (average $\pm$ standard deviation); accordingly, we required a minimum distance of about 3 s.d. (= 350 μ m) between aggregates of yellow cells to designate them as independent; the minimum number of cells required per cluster is 1.

Differentiation potential of embryonic TECs. To determine the long-term repopulation potential of embryonic TECs, isolated thymic lobes from E12.5 embryos of the fully recombined *Rosa26*-eYFP strain were dispersed into single cell suspensions¹² and stained with antibodies against CD45 and MHC class II. MHC-class-II-positive, CD45-negative, eYFP-positive cells were sorted into wells of microtitre plates that contained 2×10^5 cells from dissociated E14.5 thymic lobes of BALB/c wild-type mice. The cell suspensions were reaggregated for 2 days and then transplanted under the kidney capsule of BALB/c nude mice. After 5 months, recipient mice were killed and grafts examined for the presence of eYFP-positive cells in the ectopic thymic tissue.

Received 17 January; accepted 26 April 2006.

- Boehm, T., Bleul, C. C. & Schorpp, M. Genetic dissection of thymus development in mouse and zebrafish. *Immunol. Rev.* **195**, 15–27 (2003).
- Anderson, G. & Jenkinson, E. J. Lymphostromal interactions in thymic development and function. *Nature Rev. Immunol.* **1**, 31–40 (2001).
- Gray, D. H. *et al.* Controlling the thymic microenvironment. *Curr. Opin. Immunol.* **17**, 137–143 (2005).
- Nehls, M. *et al.* Two genetically separable steps in the differentiation of thymic epithelium. *Science* **272**, 886–889 (1996).
- Anderson, M. S. *et al.* Projection of an immunological self shadow within the thymus by the aire protein. *Science* **298**, 1395–1401 (2002); published online 10 October 2002 (doi: 10.1126/science.1075958).
- Hogquist, K. A. & Bevan, M. J. The nature of the peptide/MHC ligand involved in positive selection. *Semin. Immunol.* **8**, 63–68 (1996).
- Kyewski, B., & Derbinski, J. Self-representation in the thymus: an extended view. *Nature Rev. Immunol.* **4**, 688–698 (2004).
- Blackburn, C. C. *et al.* One for all and all for one: thymic epithelial stem cells and regeneration. *Trends Immunol.* **23**, 391–395 (2002).
- Nehls, M., Pfeifer, D., Schorpp, M., Hedrich, H. & Boehm, T. New member of the winged-helix protein family disrupted in mouse and rat nude mutations. *Nature* **372**, 103–107 (1994).

10. Blackburn, C. C. *et al.* The nu gene acts cell-autonomously and is required for differentiation of thymic epithelial progenitors. *Proc. Natl Acad. Sci. USA* **93**, 5742–5746 (1996).
11. Cunningham-Rundles, C. & Ponda, P. P. Molecular defects in T- and B-cell primary immunodeficiency diseases. *Nature Rev. Immunol.* **5**, 880–892 (2005).
12. Boehm, T., Scheu, S., Pfeffer, K. & Bleul, C. C. Thymic medullary epithelial cell differentiation, thymocyte emigration, and the control of autoimmunity require lympho-epithelial cross talk via LTbetaR. *J. Exp. Med.* **198**, 757–769 (2003).
13. Akiyama, T. *et al.* Dependence of self-tolerance on TRAF6-directed development of thymic stroma. *Science* **308**, 248–251 (2005); published online 10 February 2005 (doi: 10.1126/science.1105677).
14. Bennett, A. R. *et al.* Identification and characterization of thymic epithelial progenitor cells. *Immunity* **16**, 803–814 (2002).
15. Gill, J., Malin, M., Hollander, G. A. & Boyd, R. Generation of a complete thymic microenvironment by MTS24(+) thymic epithelial cells. *Nature Immunol.* **3**, 635–642 (2002); advance online publication 17 June 2002 (doi:10.1038/ni812).
16. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001); published online 27 March 2001 (doi: 10.1186/1471-213x-1-4).
17. Li, M. *et al.* Skin abnormalities generated by temporally controlled RXRalpha mutations in mouse epidermis. *Nature* **407**, 633–636 (2000).
18. Vassar, R., Rosenberg, M., Ross, S., Tyner, A. & Fuchs, E. Tissue-specific and differentiation-specific expression of a human K14 keratin gene in transgenic mice. *Proc. Natl Acad. Sci. USA* **86**, 1563–1567 (1989).
19. Rodewald, H. R., Paul, S., Haller, C., Bluethmann, H. & Blum, C. Thymus medulla consisting of epithelial islets each derived from a single progenitor. *Nature* **414**, 763–768 (2001).
20. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).
21. Huth, M., Schlake, T. & Boehm, T. Transposon-induced splicing defect in the rat nude gene. *Immunogenetics* **45**, 282–283 (1997).
22. Adriani, M. *et al.* Ancestral founder mutation of the nude (FOXP1) gene in congenital severe combined immunodeficiency associated with alopecia in southern Italy population. *Ann. Hum. Genet.* **68**, 265–268 (2004).
23. Frank, J. *et al.* Exposing the human nude phenotype. *Nature* **398**, 473–474 (1999).
24. Pignata, C. *et al.* Human equivalent of the mouse Nude/SCID phenotype: long-term evaluation of immunologic reconstitution after bone marrow transplantation. *Blood* **97**, 880–885 (2001).
25. Rossi, S. W., Jenkinson, W. E., Anderson, G. & Jenkinson, E. J. Clonal analysis reveals a common progenitor for thymic cortical and medullary epithelium. *Nature* doi:10.1038/nature04813 (this issue).
26. Schwenk, F., Baron, U., & Rajewsky, K. A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**, 5080–5081 (1995).
27. Nehls, M., Messerle, M., Sirulnik, A., Smith, A. J. & Boehm, T. Two large insert vectors, lambda PS and lambda KO, facilitate rapid mapping and targeted disruption of mammalian genes. *Biotechniques* **17**, 770–775 (1994).
28. Nehls, M., Pfeifer, D., Micklem, G., Schmoor, C. & Boehm, T. The sequence complexity of exons trapped from the mouse genome. *Curr. Biol.* **4**, 983–989 (1994).
29. te Riele, H., Maandag, E. R., Clarke, A., Hooper, M. & Berns, A. Consecutive inactivation of both alleles of the *pim-1* proto-oncogene by homologous recombination in embryonic stem cells. *Nature* **348**, 649–651 (1990).
30. Schorpp, M., Hofmann, M., Dear, T. N. & Boehm, T. Characterization of mouse and human nude genes. *Immunogenetics* **46**, 509–515 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Maul-Pavicic, S. Groß, C. Sainz-Rueda, A. Haas-Assenbaum, C. Happe, E. Nikolopoulos and M. Konrath for help during various stages of this project, and K. Rajewsky, P. Soriano, P. Chambon, S. Srinivas, K. Eichmann, G. Turchinovich, for mouse lines or reagents. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Max-Planck Society.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.B. (boehm@immunbio.mpg.de).

Nanog promotes transfer of pluripotency after cell fusion

José Silva¹, Ian Chambers¹, Steven Pollard¹ & Austin Smith^{1,2}

Through cell fusion, embryonic stem (ES) cells can erase the developmental programming of differentiated cell nuclei and impose pluripotency^{1,2}. Molecules that mediate this conversion should be identifiable in ES cells. One candidate is the variant homeodomain protein Nanog, which has the capacity to entrain undifferentiated ES cell propagation^{3,4}. Here we report that in fusions between ES cells and neural stem (NS) cells, increased levels of Nanog stimulate pluripotent gene activation from the somatic cell genome and enable an up to 200-fold increase in the recovery of hybrid colonies, all of which show ES cell characteristics. Nanog also improves hybrid yield when thymocytes or fibroblasts are fused to ES cells; however, fewer colonies are obtained than from ES×NS cell fusions, consistent with a hierarchical susceptibility to reprogramming among somatic cell types. Notably, for NS×ES cell fusions elevated Nanog enables primary hybrids to develop into ES cell colonies with identical frequency to homotypic ES×ES fusion products. This means that in hybrids, increased Nanog is sufficient for the NS cell epigenome to be reset completely to a state of pluripotency. We conclude that Nanog can orchestrate ES cell machinery to instate pluripotency with an efficiency of up to 100% depending on the differentiation status of the somatic cell.

A differentiated epigenome can be reverted to pluripotency by nuclear transfer into enucleated oocytes⁵ or by cell fusion with pluripotent stem cells^{1,6–11}. It is speculated that the epigenome of tissue stem cells may be more readily converted than that of more terminally restricted cells, but this has been hard to test owing to the difficulty in obtaining pure populations of tissue stem cells. However, neural stem (NS) cells with the potency to differentiate into neurons, astrocytes and oligodendrocytes can now be stably propagated as seemingly homogeneous populations¹². NS cells can be derived from the *in vitro* differentiation of embryonic stem (ES) cells, or from fetal or adult mouse brain, and in all cases show similar messenger RNA and protein expression profiles that are distinctive from ES cells¹³.

Molecules that mediate epigenetic deprogramming and transcriptional resetting^{14,15} are not known, but must be present in early embryos¹⁶ and pluripotent ES cells. Nanog is a divergent homeodomain protein expressed in early embryos with the capacity to sustain cell-autonomous pluripotency in ES cells^{3,4,17}. We investigated whether Nanog might facilitate transfer of pluripotency to NS cell nuclei in the context of cell fusion (Supplementary Fig. 1).

The *Oct4*-GFPirespac (O4G) transgene, in which enhanced green fluorescent protein (eGFP) is expressed under the control of an *Oct4* (*Pou5f1*) regulatory element and is active only in pluripotent and germline cells, can be exploited as a convenient indicator of acquisition of pluripotency^{2,18}. We used poly(ethylene glycol) (PEG)-induced fusion⁸ to generate cell hybrids between NS-O4G cells and ES cells. Parental RH ES cells constitutively express dsRed2 and hygromycin resistance. They were fused with NS-O4G cells and hybrids selected in

hygromycin plus puromycin. Drug-resistant colonies consisting of cells that expressed both red and green fluorescence were obtained at a frequency of 1.4×10^{-6} . RHN ES cells are a transfected derivative of RH cells that stably overexpress Nanog. When fused to NS-O4G cells, the yield of double-resistant dual-fluorescent colonies was dramatically increased. At least 200-fold more colonies were obtained compared with parallel RH×NS-O4G fusions (Fig. 1a). To confirm the reproducibility and specificity of this effect of Nanog overexpression, we carried out fusions between NS cells and either EF4 ES cells that express a floxed *Nanog* transgene, or their derivative EF4Cre3 from which the transgene has been excised by Cre recombinase³. In agreement with the observations from RHN and RH fusions, EF4×NS cell fusions produced abundant drug-resistant colonies on each plate, whereas EF4Cre3×NS cell fusions yielded no colonies from equivalent numbers of cells plated (Fig. 1b). To exclude idiosyncratic behaviour of NS-O4G cells, we repeated the fusions with independent NS cell lines. NS-TGFP is a fetal-brain-derived line that constitutively expresses the fusion protein Tau-GFP and puromycin resistance¹⁹. RHN×NS-TGFP fusions generated numerous hybrid colonies, whereas RH×NS-TGFP fusions produced only a single colony (Fig. 1c). This difference was also observed for fusions of RHN and RH ES cells with ES-cell-derived and adult-brain-derived NS cells (Supplementary Fig. 2). To confirm correlation with Nanog overexpression, we compared the levels of Nanog, Oct4 and Sox2 proteins between the ES cell lines (Fig. 1e). Similar elevated levels of Nanog were detected in RHN and EF4 ES cells, whereas 4–5-fold lower amounts were present in the parental and revertant ES cell lines²⁰. In control ES×ES cell fusions, *Nanog* transgene expression resulted in a less than twofold increase in hybrid yield (Fig. 1d). These results show that Nanog overexpression in ES cells enhances significantly the ability to generate viable hybrids from fusions with NS cells.

The hybrid colonies displayed compact morphology and histochemical staining characteristic of ES cells, albeit individual cells had the enlarged nuclei expected for tetraploid hybrids. When fusion products were selected in NS cell medium, which does not support ES cell growth, no colonies formed. These observations suggest that reprogramming is a unidirectional process that either goes to completion, resulting in an ES cell phenotype, or generates non-viable cell states (Supplementary Fig. 1). To define the status of ES×NS cell hybrids, we examined chromosomal content, epigenetic marks and gene expression. Chromosome counts for 12 independent hybrid clones originating from fusions of RH or RHN ES cells with NS cells all gave a modal number of 80 chromosomes (Supplementary Fig. 3). Consistent with their somatic status, XX NS cells have an inactive X chromosome marked by nuclear bodies of histone H3 trimethylated on lysine 27 (H3K27me3)²¹, monoubiquitylated histone H2A (H2Aub1)²², and *Xist* RNA (Fig. 2a–c). Deprogramming in ES×NS cell hybrids was evident by the disappearance of the

¹Centre Development in Stem Cell Biology, Institute for Stem Cell Research, University of Edinburgh, Edinburgh, EH9 3JQ, UK. ²Institute for Stem Cell Biology University of Cambridge, Cambridge, CB2 1QT, UK.

repressive histone marks (Fig. 2a, b), and replacement of the *Xist* RNA nuclear body by a pinpoint signal corresponding to nascent transcripts characteristic of ES cells²³ (Fig. 2c). We then examined allelic expression of *Oct4* by fluorescence *in situ* hybridization (FISH). No signal was observed in NS cells, whereas hybrids showed three or four pinpoint signals per nucleus demonstrating reactivation of the *Oct4* alleles of NS-cell origin (Fig. 2d). Conversely, RT-PCR (polymerase chain reaction following reverse transcription) analyses revealed that *Olig2* and *Blbp* (*Fabp7*), which are expressed in NS cells but not ES cells¹², are silenced in ES×NS hybrid cells (Fig. 2e). Finally, to determine whether the hybrids acquire non-neural differentiation competence, the *Nanog* overexpression cassette was excised through *cre* transfection, releasing them for differentiation³ (Supplementary Fig. 4). Embryoid bodies were analysed for expression of the early mesoderm marker *T-brachyury* and the endoderm marker *Gata4* (Fig. 2f). Neither is detectable in the NS cells nor in undifferentiated hybrids. However, expression of both genes is upregulated during embryoid-body differentiation. Embryoid bodies were then allowed to outgrow. Most (24 out of 35) wells showed beating cells and stained for the myogenic marker α -actinin (Supplementary

Fig. 4e). Collectively, these data show that the NS cell genome is reprogrammed to an ES cell identity, and that the hybrids possess multilineage differentiation capacity.

Cell-cycle phase and population doubling time of ES cells are unchanged by *Nanog* transgene expression (data not shown). *Nanog* does reduce the small fraction (<5%; Supplementary Fig. 5a) of differentiating cells normally present³. However, purification of undifferentiated ES cells^{24,25} before fusion did not abolish the higher hybrid yields obtained from *Nanog*-expressing ES cells (Supplementary Fig. 2). To evaluate whether *Nanog* might increase hybrid yield through enhanced cell viability and clonogenicity, we compared the effect of PEG treatment on RH and RHN ES cells. Cell survival—scored as the number of viable cells 24 h after PEG treatment—was indistinguishable for RH and RHN ES cells (Supplementary Fig. 5). Cells were then deposited into microwells and the formation of undifferentiated ES cell colonies scored by alkaline phosphatase staining after 12 days. PEG treatment did not noticeably increase differentiation for either cell line. However, the total number of colonies was reduced after PEG treatment, and this effect was more marked for RH, rather than RHN, cells (Supplementary Fig. 5). These data indicate that *Nanog* overexpression enhances resistance to PEG treatment, resulting in a 1.5–2-fold increase in colony number. This is in line with the effect on ES×ES cell hybrid colony formation (Fig. 1e), but is insufficient to account for the observed >100-fold increase in ES×NS hybrid colony yield.

Another explanation for the effect of *Nanog* overexpression would be if this rendered ES cells more 'fusogenic'. Therefore, fusions between RH×NS-TGFP and RHN×NS-TGFP were analysed by fluorescence-activated cell sorting (FACS) after 24 h to determine the frequency of fused products detected by dual red and green fluorescence. As expected, given the low overall efficiency of PEG fusion, the number of genuine double-positive cells is relatively low. Importantly, however, the frequency did not differ appreciably between the two groups, with RH and RHN fusions containing $0.19 \pm 0.025\%$ and $0.14 \pm 0.04\%$ (average $\pm$ s.d.) double-positive cells, respectively (Fig. 3a–c). Inspection of cytopins showed that 90% or more of dual-fluorescent cells are mononucleate, indicating that they have progressed through the heterokaryon stage (Supplementary Fig. 6). The proportion of mononucleate cells is not affected by overexpression of *Nanog*. Therefore, *Nanog* overexpression does not increase the frequency of fusion or the formation of primary hybrids.

We then investigated whether *Nanog* alters the frequency of primary hybrids that activate the pluripotent marker *Oct4*-GFP from the NS cell genome. RH×NS-O4G or RHN×NS-O4G cell fusion products were examined by FACS for expression of GFP. Consistent with previous reports^{1,26}, *Oct4*-GFP production from the somatic genome is first detected 48 h after fusion. Notably, however, at least 40-fold-more GFP-positive cells are present in RHN, rather than RH, fusions (Supplementary Fig. 7), establishing that the primary effect of *Nanog* is independent of hybrid selection. FACS purification was then used to determine the colony-forming efficiency of fused cells. This will depend both on productive resolution of epigenome differences and on cell-cycle coordination between the two genomes. In two independent experiments, the generation of colonies from sorted primary hybrids from RH×NS-TGFP fusions was below the limit of detection (<1 in 1,600) (Fig. 3d, e). In contrast, RHN×NS-TGFP fusions produced hybrid colonies at a frequency >3% of mononucleate hybrids plated (Fig. 3d, e). These data indicate that *Nanog* acts in primary hybrids to promote the conversion of the NS cell epigenome to a pluripotent state.

To investigate whether the effect of *Nanog* depends on a specific predisposition of NS cells, or is generalizable to other somatic cell fusion partners, we carried out fusions of RH and RHN cells with primary cultures of mouse embryonic fibroblasts (MEFs) and adult mouse thymocytes. In both cases, the yield of hybrid colonies was at least an order of magnitude greater from RHN, rather than RH, fusions (Supplementary Fig. 2), although lower than for NS cell

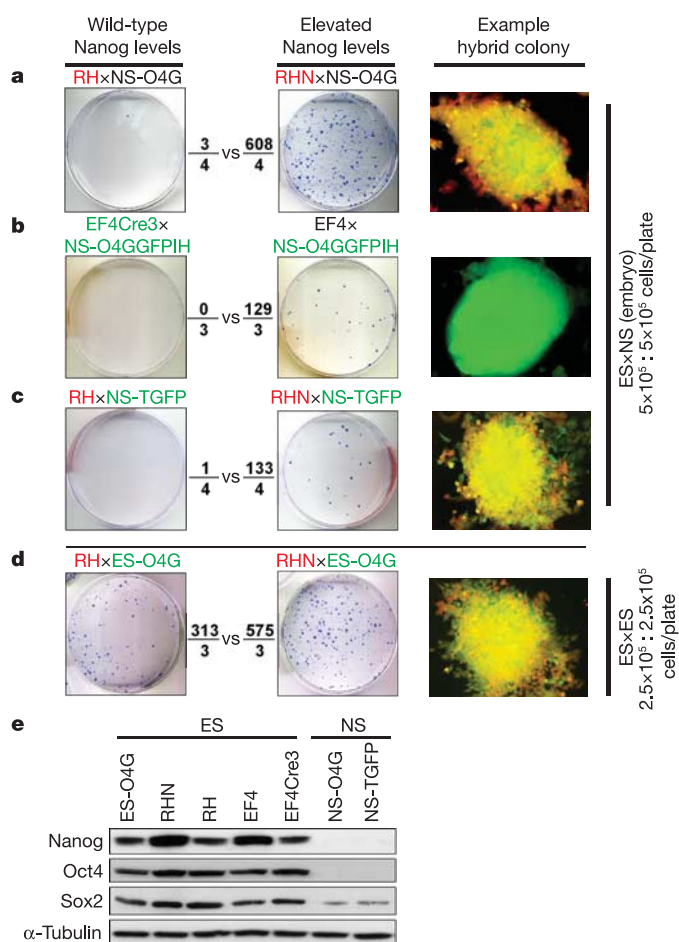


Figure 1 | Elevated levels of *Nanog* enhances formation of hybrid colonies from ES×NS cell fusions. **a–c**, Plates containing hybrid colonies from PEG fusions of ES cells and NS cells (left and middle panels). The number of hybrid colonies/plates is shown beside each panel. Fluorescence images of example hybrid colonies are also shown (right panels). **d**, Control homotypic fusions of RH and RHN cells with ES-O4G cells. **e**, Immunoblotting analysis of the levels of *Nanog*, *Oct4* and *Sox2* in ES cells and NS cells. α -Tubulin was used as a loading control. Note that ES-O4G cells were grown under puromycin selection for expression of the *Oct4* transgene to generate a relatively pure undifferentiated ES cell population²⁵. Refer to the text and Methods for an explanation of the cell lines used.

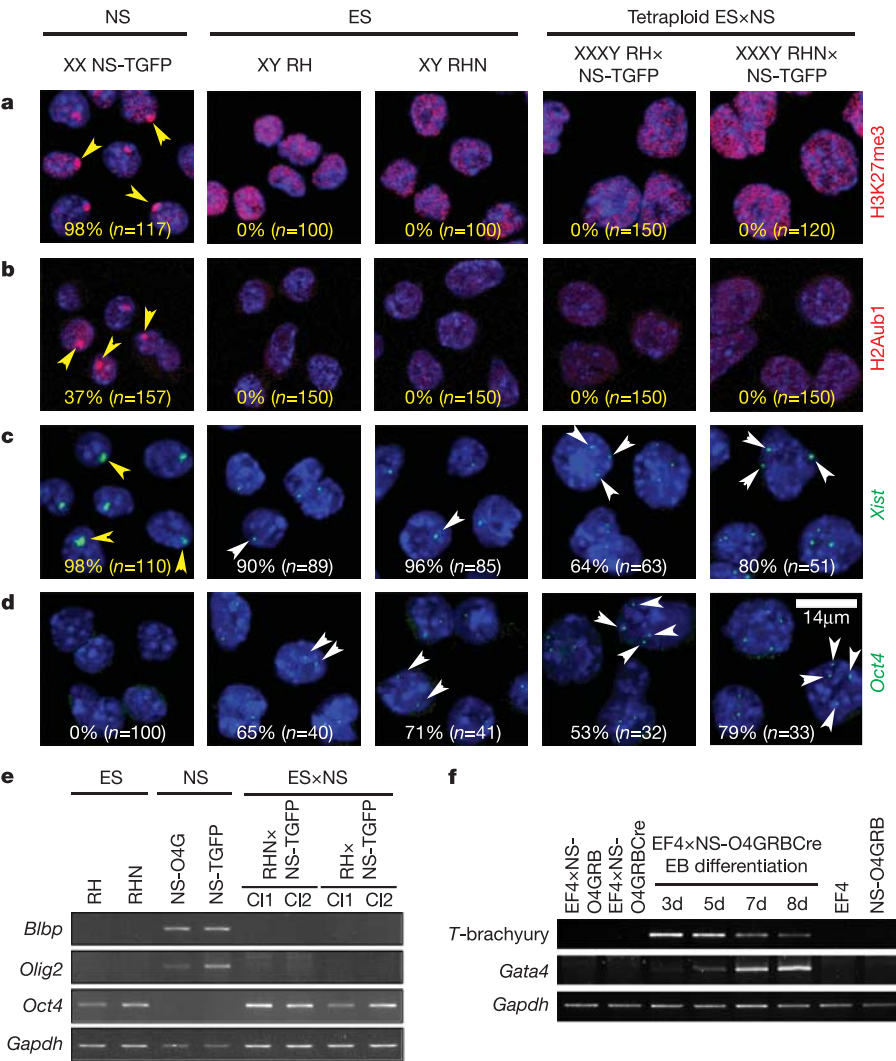


Figure 2 | Characterization of ES x NS cell hybrids. **a–d**, Immunofluorescence staining for H3K27me3 (**a**) and H2Aub1 (**b**), and RNA FISH staining for *Xist* (**c**) and *Oct4* (**d**), in XX NS cells, XY ES cells, and XXXY hybrids. Yellow arrowheads indicate the presence of nuclear bodies in NS-TGFP cells. The percentage of cells displaying these markers is indicated in yellow. White arrowheads indicate *Oct4* or *Xist* RNA nascent transcripts. The percentage of cells showing expected allelic expression is indicated. All nuclei were counterstained with DAPI (4,6-diamidino-2-phenylindole). **e**, RT-PCR analysis of *Blbp*, *Olig2* and *Oct4* expression. **f**, RT-PCR analysis of *T-brachyury* and *Gata4* expression in EF4 x NS-O4GRB hybrids after Cre excision of the *Nanog* transgene and induction of embryoid body differentiation for the indicated number of days. *Gapdh* was used as a ubiquitously expressed control. Cl, clone; EB, embryoid body.

fusions. To control for differences in fusogenicity we purified the primary fusion products. Analysis of the FACS-sorted populations confirmed that Nanog did not alter the frequency of fusion. More than 90% of true double-positive cells were mononucleate hybrids for MEF fusions, but a subpopulation of around 36% of fused products were binucleate in the thymocyte fusions (Supplementary Fig. 6). RH primary hybrids failed to produce colonies from a total of 1.2×10^4 and 5×10^3 cells plated for MEF and thymocyte fusions, respectively. RHN primary hybrids, in contrast, did produce colonies, confirming that overexpression of Nanog promotes the transfer of an ES cell phenotype to different types of somatic nuclei. However, hybrid

colonies develop at a lower efficiency from MEF fusions than from thymocyte primary fusion products (0.08% compared with 0.33%) (Fig. 3a–e). This difference is increased if the heterokaryons in the thymocyte fusions are excluded. Nonetheless, RHN x thymocyte fusion products still yield 5–10-fold fewer hybrid colonies than RHN x NS fusions (Table 1). Although cell-cycle differences are likely to be a major factor affecting hybrid viability independently of reprogramming, the data are also consistent with the developmental status of the somatic cell partner being a critical parameter in phenotype conversion. Interestingly, after nuclear transfer into mouse oocytes, NS cell nuclei seem to give rise to ES cells more

Table 1 | ES x NS primary hybrids form pluripotent colonies with similar efficiency to ES x ES primary hybrids

Expt	Hybrid pair	No. FACS-sorted cells	FACS purity	No. hybrid colonies	Colonies/plated hybrids (%) [*]	Colonies/mononucleate hybrids (%) [†]
1	RHN x NS-TGFP	2,000	40%	34	4.25	4.59
2	RHN x NS-TGFP	700	40%	11	3.93	4.25
1	RHN x ES-O4G	4,500	70%	106	3.37	3.66
2	RHN x ES-O4G	700	80%	22	3.93	4.27
1	RHN x MEF-TGFP	10,500	70%	6	0.08	0.09
2	RHN x MEF-TGFP	1,500	70%	0	0	0
1	RHN x T-TGFP	4,400	90%	13	0.33	0.51
2	RHN x T-TGFP	1,200	90%	4	0.35	0.58

^{*}Percentage of hybrid colonies from plating FACS-purified primary fused products from the indicated fusions. Scores are corrected for purity of the sorted hybrids as determined by FACS analysis.
[†]Scores adjusted for nuclear content of primary fused products.
Expt, experiment number.

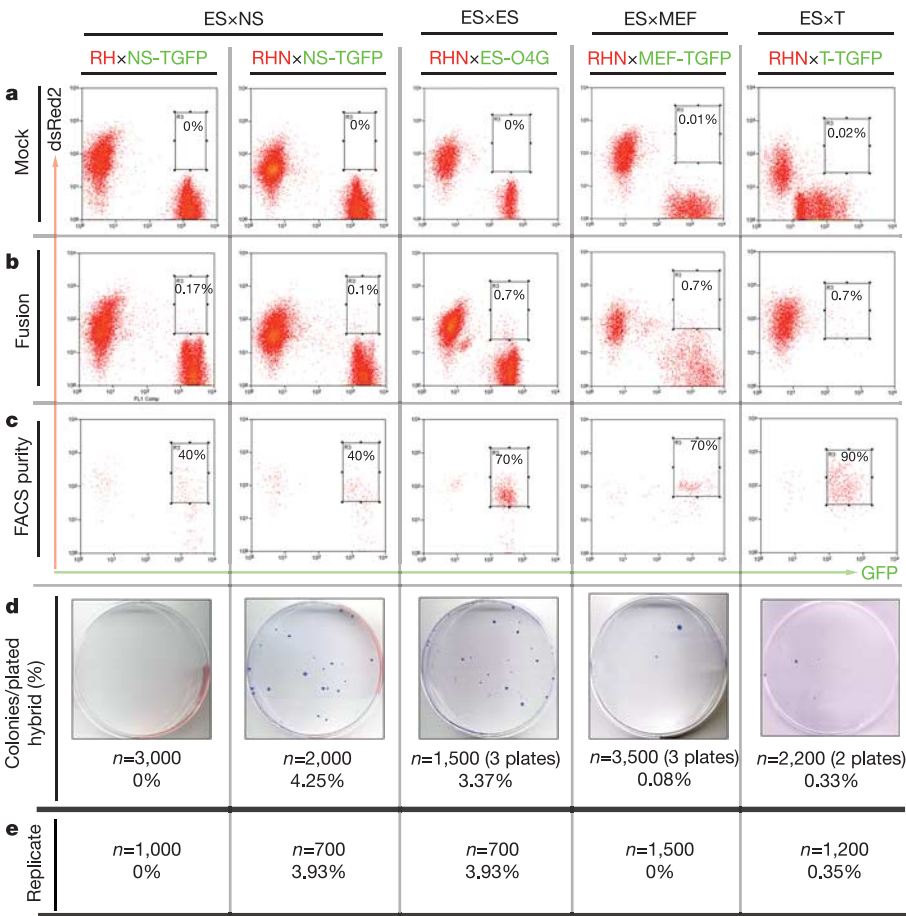


Figure 3 | FACS purification and analysis of primary fusion products. **a–c**, FACS analysis of both red and green fluorescence. **a**, Mock fusion providing a gate control (boxed) for the analysis and FACS-sorting of the hybrid population. **b**, Fusion mixture harvested 24 h after PEG treatment (non-adherent unfused thymocytes are excluded). **c**, Sorted populations gated in **b** subjected to FACS analysis to determine the proportion of true double-positive hybrids. **d**, Primary hybrids sorted in **b** were plated and the number of colonies formed was scored. Percentage scores are adjusted for purity of the FACS-sorted cells. **e**, Scores from a replicate experiment are indicated.

readily than nuclei of other differentiated cell types and at an efficiency comparable to ES cell nuclei²⁷. It remains to be determined whether the advantage of NS cells resides in their stem-cell nature or in some other feature, such as their expression of Sox2—a major player in the transcriptional network of pluripotency^{17,28}.

We examined how the frequency of conversion of NS cells to pluripotency in the presence of elevated Nanog compares with homotypic ESxES fusions in which there is no requirement for deprogramming and resetting. RHN ES cells were fused with ES cells carrying the O4G transgene, and in parallel fused with NS-TGFP cells. Remarkably, the frequency of hybrid colonies recovered from double-positive FACS-sorted primary hybrids was not significantly different between the two fusions (Table 1 and Fig. 3d, e). We conclude that, in the presence of high levels of Nanog, developmental differences between ES cells and NS cells are extinguished in hybrids. These data also reveal that the vast majority (96%) of primary hybrids fail to produce colonies for reasons seemingly unrelated to reprogramming as they arise in ESxES fusions. Such factors may include cell-cycle asynchrony, collateral damage induced by the fusion process, and the relatively low cloning efficiency of ES cells.

The level of Nanog emerges as the limiting factor for converting the NS cell epigenome to pluripotency in hybrids. This raises the question of whether Nanog alone is sufficient to convert NS cells to ES cells. Therefore, we transfected NS-O4G cells with a *Nanog* expression vector and obtained NS-O4GN cells that stably express the Nanog protein at a similar level to ES cells (Fig. 4a). NS-O4GN cells show no overt change in phenotype, proliferate normally, and retain expression of the NS cell markers *Blbp* and *Olig2* (Fig. 4b, c). When plated into standard ES cell culture medium, containing leukaemia inhibitory factor (LIF) plus serum or bone morphogenetic protein (BMP), NS-O4GN cells differentiate into astrocytic cells just as wild-type NS cells do¹². There is no activation of the *Oct4*-GFP

transgene or of endogenous *Oct4*, even on treatment with trichostatin A and 5-azacytidine—two powerful epigenetic destabilizers (Supplementary Fig. 8). Thus, Nanog seems to be inert in NS cells. Nonetheless, when NS-O4GN cells were fused to RH ES cells, the hybrid colony yield was increased eightfold compared with O4G fusions (Fig. 4d). This effect is smaller than when Nanog is delivered

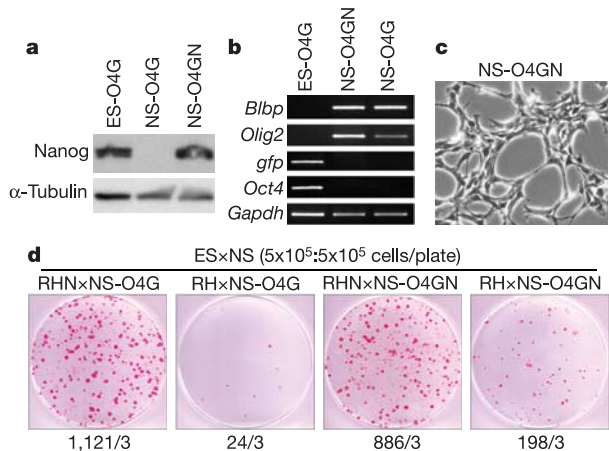


Figure 4 | Nanog expression in NS cells increases the frequency of hybrid colonies. **a**, Immunoblot showing that NS-O4GN cells express Nanog protein in at least comparable amounts to ES-O4G cells. **b**, RT-PCR analysis showing that NS-O4GN cells retain expression of *Blbp* and *Olig2*, with no detectable activation of *Oct4*-GFP or *Oct4* expression. **c**, NS-O4GN cells show typical NS cell morphology. **d**, Plates containing hybrid colonies from the indicated fusions stained with alkaline phosphatase. The number of hybrid colonies/plates is shown under each panel.

from the ES cell, which may simply reflect a lower overall amount of Nanog compared with overexpression in ES cells. Alternatively, there may be a narrow time-window for the post-translational modification of Nanog²⁰, formation of functional protein complexes, or induction of target genes after fusion. Clearly, however, enhanced transfer of pluripotency does not require prior alteration of the ES cell state. Although we cannot rule out completely some form of preconditioning of either ES cells or NS cells before fusion, these data suggest that the locus of Nanog action is the primary hybrid.

Together, these findings demonstrate that Nanog has a dominant role in instating pluripotency in an ES cell hybrid. However, Nanog does not operate alone to transfer pluripotency but acts in conjunction with additional ES cell machinery. Nanog may act as a master transcriptional organizer that entrains the hierarchy of pluripotent gene expression after erasure of the differentiated epigenome¹⁵. Alternatively, or additionally, Nanog may orchestrate the activity of yet-to-be-defined chromatin remodelling factors. Identifying those proteins that interact with Nanog is now crucial to illuminate its mode of action. This may ultimately enable strategies to be designed for converting a somatic cell to a pluripotent state without recourse to nuclear transfer or cell fusion.

METHODS

Cells. NS cells were obtained from HP165 mice, carrying regulatory sequences of the mouse *Oct4* gene¹⁸ driving GFP and puromycin resistance², and TP3 mice, expressing constitutively the fusion protein Tau-GFP (TGFP) and the puromycin-resistance gene¹⁹. Thymocytes (T-TGFP) were obtained from 6–8-week-old mice. Mouse embryo fibroblasts (MEF-TGFP) were derived from 13.5-dpc (days post-coitum) embryos. NS cells were derived from 14.5-dpc fetal forebrain (NS-O4G and NS-TGFP), adult lateral ventricle (ANS-O4G) and ES-O4G cells as described elsewhere^{12,13}. NS-O4G stable transfectants were generated by electroporation; NS-O4GRB cells express the dsRed2 fluorescent protein and blasticidin resistance; NS-O4GN cells express Nanog and puromycin resistance; NS-O4GNB cells express Nanog and blasticidin resistance; and NS-O4GGFPIH cells express GFP and hygromycin resistance. Culture and differentiation conditions of ES cells, and generation of EF4 and EF4Cre3 ES cells, have been described elsewhere³. RH cells were derived from E14tg2a by transfection of pHPGAGdsRed2, and express constitutively dsRed2 and hygromycin resistance. RH cells were additionally transfected with pPyCAGNanogIZ to obtain RHN cells that express constitutively Nanog and zeocin resistance.

Cell hybrids. For PEG fusions, cells of each type (generally 4×10^6), were combined in serum-free GMEM in a conical tube, pelleted, and the supernatant aspirated. The pellet was broken by gentle tapping, and 300 μ l of 50% w/v PEG 1500 (Roche) prewarmed to 37°C was gently added. Cells were left in the 50% PEG solution over a 3-min period with occasional stirring. Then, 1 ml of medium was added over a period of 3 min. Subsequently, an additional 3 ml of medium was added. Cells were spun down and the supernatant discarded. The pellet was resuspended in complete ES cell medium and plated. Selection was applied after 24 h using hygromycin (250 μ g ml⁻¹), puromycin (1 μ g ml⁻¹), zeocin (100 μ g ml⁻¹) and blasticidin (10 μ g ml⁻¹) as appropriate. As EF4 cells are puromycin resistant, hybrids are selected in hygromycin after fusion with NS-O4GGFPIH cells. The Nanog transgene in EF4 $\times$ NS-O4GRB hybrids was removed by transient transfection with pPyCAGcreIP using Lipofectamine 2000 (Invitrogen).

Received 26 January; accepted 16 May 2006.

Published online 14 June 2006.

1. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. & Tada, T. Nuclear reprogramming of somatic cells by *in vitro* hybridization with ES cells. *Curr. Biol.* **11**, 1553–1558 (2001).
2. Ying, Q. L., Nichols, J., Evans, E. P. & Smith, A. G. Changing potency by spontaneous fusion. *Nature* **416**, 545–548 (2002).
3. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
4. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).

5. Campbell, K. H. S., McWhir, J., Ritchie, W. A. & Wilmut, I. Sheep cloned by nuclear transfer from a cultured cell line. *Nature* **380**, 64–66 (1996).
6. Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369–1373 (2005).
7. Kimura, H., Tada, M., Nakatsuji, N. & Tada, T. Histone code modifications on pluripotential nuclei of reprogrammed somatic cells. *Mol. Cell. Biol.* **24**, 5710–5720 (2004).
8. Matveeva, N. M. *et al.* *In vitro* and *in vivo* study of pluripotency in intraspecific hybrid cells obtained by fusion of murine embryonic stem cells with splenocytes. *Mol. Reprod. Dev.* **50**, 128–138 (1998).
9. Miller, R. A. & Ruddle, F. H. Teratocarcinoma $\times$ friend erythroleukemia cell hybrids resemble their pluripotent embryonal carcinoma parent. *Dev. Biol.* **56**, 157–173 (1977).
10. Tada, M., Tada, T., Lefebvre, L., Barton, S. C. & Surani, M. A. Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J.* **16**, 6510–6520 (1997).
11. Miller, R. A. & Ruddle, F. H. Pluripotent teratocarcinoma–thymus somatic cell hybrids. *Cell* **9**, 45–55 (1976).
12. Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3**, 1596–1606 (2005).
13. Pollard, S. M., Conti, L., Sun, Y., Goffredo, D. & Smith, A. Adherent neural stem (NS) cells from foetal and adult forebrain. *Cerebral Cortex* (in the press).
14. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917–929 (2005).
15. Smith, A. The battlefield of pluripotency. *Cell* **123**, 757–760 (2005).
16. Bao, S. *et al.* Initiation of epigenetic reprogramming of the X chromosome in somatic nuclei transplanted to a mouse oocyte. *EMBO Rep.* **6**, 748–754 (2005).
17. Chambers, I. & Smith, A. Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23**, 7150–7160 (2004).
18. Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonic cells. *Development* **122**, 881–894 (1996).
19. Pratt, T., Sharp, L., Nichols, J., Price, D. J. & Mason, J. O. Embryonic stem cells and transgenic mice ubiquitously expressing a tau-tagged green fluorescent protein. *Dev. Biol.* **228**, 19–28 (2000).
20. Yates, A. & Chambers, I. The homeodomain protein Nanog and pluripotency in mouse embryonic stem cells. *Biochem. Soc. Trans.* **33**, 1518–1521 (2005).
21. Silva, J. *et al.* Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed–Enx1 polycomb group complexes. *Dev. Cell* **4**, 481–495 (2003).
22. de Napoles, M. *et al.* Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* **7**, 663–676 (2004).
23. Sheardown, S. A. *et al.* Stabilization of Xist RNA mediates initiation of X chromosome inactivation. *Cell* **91**, 99–107 (1997).
24. Mountford, P. *et al.* Dicistronic targeting constructs: reporters and modifiers of mammalian gene expression. *Proc. Natl Acad. Sci. USA* **91**, 4303–4307 (1994).
25. Mountford, P., Nichols, J., Zevnik, B., O'Brien, C. & Smith, A. Maintenance of pluripotential embryonic stem cells by stem cell selection. *Reprod. Fertil. Dev.* **10**, 527–533 (1998).
26. Do, J. T. & Scholer, H. R. Nuclei of embryonic stem cells reprogram somatic cells. *Stem Cells* **22**, 941–949 (2004).
27. Belloch, R. *et al.* Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. *Stem Cells* doi:10.1634/stemcells.2006-0050 (18 May 2006).
28. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Vrana for assistance with FACS and Y. Costa for critical reading of the manuscript. This research was supported by the Biotechnological and Biological Sciences Research Council and the Medical Research Council of the United Kingdom, the Wellcome Trust, and by an EMBO Long-term Fellowship (J.S.).

Author Contributions J.S. designed and executed experiments, analysed data, and drafted the paper; I.C. and S.P. generated reagents and contributed to experimental design; and A.S. designed experiments and wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.S. (austin.smith@ed.ac.uk).

LETTERS

Golgi maturation visualized in living yeast

Eugene Losev¹, Catherine A. Reinke¹, Jennifer Jellen¹, Daniel E. Strongin¹, Brooke J. Bevis¹ & Benjamin S. Glick¹

The Golgi apparatus is composed of biochemically distinct early (*cis*, *medial*) and late (*trans*, TGN) cisternae. There is debate about the nature of these cisternae^{1–3}. The stable compartments model predicts that each cisterna is a long-lived structure that retains a characteristic set of Golgi-resident proteins. In this view, secretory cargo proteins are transported by vesicles from one cisterna to the next. The cisternal maturation model predicts that each cisterna is a transient structure that matures from early to late by acquiring and then losing specific Golgi-resident proteins. In this view, secretory cargo proteins traverse the Golgi by remaining within the maturing cisternae. Various observations have been interpreted as supporting one or the other mechanism^{4–9}. Here we provide a direct test of the two models using three-dimensional time-lapse fluorescence microscopy of the yeast *Saccharomyces cerevisiae*. This approach reveals that individual cisternae mature, and do so at a consistent rate. In parallel, we used pulse–chase analysis to measure the transport of two secretory cargo proteins. The rate of cisternal maturation matches the rate of protein transport through the secretory pathway, suggesting that cisternal maturation can account for the kinetics of secretory traffic.

Our experimental strategy was to visualize individual cisternae and determine whether the composition of Golgi-resident proteins changed over time (Supplementary Fig. S1). According to the stable compartments model, if an early Golgi protein were tagged with green fluorescent protein (GFP), then an early cisterna should exhibit green fluorescence indefinitely. However, according to the cisternal maturation model, an early cisterna should acquire green fluorescence, remain fluorescent for a brief period and then lose fluorescence. A further test would be to tag an early Golgi protein with GFP and a late Golgi protein with DsRed. If Golgi compartments are stable, a given cisterna should stay the same fluorescent colour, but if the compartments are transient, each cisterna should change colour from green to red (Supplementary Movies S1a, b).

Although these experiments are conceptually simple, they are experimentally challenging because light microscopy cannot distinguish individual cisternae in a stacked Golgi. We circumvented this problem using *S. cerevisiae*, in which Golgi cisternae are not stacked and are therefore resolvable by fluorescence microscopy^{5,10,11}. Cisternal stacking probably does not affect the basic mode of Golgi operation, because the budding yeast *Pichia pastoris* has a stacked Golgi but resembles *S. cerevisiae* with regards to secretion kinetics and Golgi compartmentation^{12,13}, and because the secretory machinery is conserved between yeast and mammals¹⁴.

We examined Golgi dynamics at 23 °C using three-dimensional time-lapse confocal microscopy (that is, four-dimensional or 4D microscopy)^{15,16}. Every 1–3 s, we captured a z-stack of 16–24 optical sections spanning the entire cell. The optical sections for each z-stack were then collapsed to generate a 2D projection (Supplementary Fig. S2a). These projections were corrected for photobleaching so that total fluorescence signals remained constant.

The main hurdle was to generate fluorescently tagged Golgi marker proteins that localized correctly at high expression levels

without significantly affecting the organelle. Lumenally-oriented Golgi proteins were excluded because GFP folds poorly in the lumen of the yeast secretory pathway⁵. One suitable marker was Sec7, an abundant protein that associates peripherally with late Golgi cisternae^{10,17}. Sec7 is concentrated on about half of the cisternae in a cell¹³. The second marker was Vrg4, an abundant transmembrane protein that carries GDP-mannose into the early Golgi^{18,19}. Both proteins could be tagged by gene replacement without disrupting their essential functions. Mild (~3-fold) or strong (~12-fold) overexpression of tagged Sec7, or mild (~2-fold) overexpression of tagged Vrg4 had no apparent effect on growth rate (data not shown) or on Golgi compartmentation and secretory pathway activity (Supplementary Figs S3, S4).

When Sec7 or Vrg4 was tagged with GFP, each cell contained multiple fluorescent spots that presumably represented individual Golgi cisternae⁵. These cisternae were mobile and often difficult to follow in the 2D projections. We therefore took advantage of the full 4D data sets, which allowed us to track individual cisternae. For example, Supplementary Movie S2 shows a cisterna being tracked through 20 consecutive z-stacks, and Supplementary Fig. S2b shows two cisternae that seem to be connected in a 2D projection but are clearly separate in the corresponding optical sections. Our 4D analysis indicated that cisternal fusion and fission events were uncommon. However, most of the cisternae occasionally passed close enough to other cisternae that they could not be fully resolved even in the optical sections. This problem was exacerbated by the tendency of Golgi cisternae to cluster near sites of polarized growth^{11,17}. Each movie ultimately yielded only a few cisternae that could be resolved unambiguously at every time point. We analysed those cisternae quantitatively. As far as we could judge from extensive qualitative observation, most or all cisternae had the same properties as the ones used for quantification.

We first made movies of an *S. cerevisiae* strain expressing a Sec7–GFP fusion protein (Supplementary Movie 1a). To simplify the presentation of relevant data, we also made edited movies showing only the cisternae that were used for quantification (Supplementary Movie 1b). Individual cisternae acquired fluorescence, remained fluorescent for some time and then lost fluorescence (Fig. 1a, b). Analysis of 29 cisternae from a total of ten cells revealed that the duration of labelling with Sec7–GFP was fairly consistent, averaging ~2 min and ranging from about 1 to 3 min (Fig. 1c and Supplementary Table S1). We obtained similar results with a strain expressing GFP–Vrg4 (Supplementary Fig. S5, Supplementary Movie S3a, b and Supplementary Table S1). This transient labelling is consistent with the cisternal maturation model.

Do these marker proteins accurately reflect the dynamics of Golgi cisternae, or do different markers have distinct dynamics? To address this question, we tagged the late Golgi simultaneously with two fluorescent proteins. DsRed-Monomer was fused to Sec7, and GFP was fused to Sys1, a transmembrane protein that recruits the GTPase Arl3 (ref. 20). Sec7–DsRed and Sys1–GFP colocalized extensively (Fig. 2a). Notably, the two proteins often seemed to be partially

¹Department of Molecular Genetics and Cell Biology, and Institute for Biophysical Dynamics, The University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA.

segregated within a given cisterna (Fig. 2c). This observation might indicate the existence of functional domains within cisternae, but might also be an effect of overexpression. In any case, dual-colour 4D imaging confirmed that Sec7–DsRed and Sys1–GFP showed very similar transient labelling of cisternae (Fig. 2b). Two representative cisternae are shown in Fig. 2 and Supplementary Movie 2a, b. In further experiments, GFP was fused to the transmembrane soluble NSF attachment protein receptor (SNARE) Gos1 (refs 15, 21) or the Golgi-localized rhomboid protease Rbd2 (ref. 22). Again, we observed transient labelling of cisternae, although quantification was not performed because GFP–Gos1 overexpression slightly altered the number and size of cisternae, and Rbd2–GFP resulted in some background non-Golgi labelling (data not shown). Thus, we conclude that multiple classes of Golgi proteins share the property of labelling an individual cisterna only transiently.

On the basis of these findings, we attempted to visualize cisternal maturation by simultaneously tagging the early Golgi with GFP–Vrg4 and the late Golgi with Sec7–DsRed. Although the

resulting fluorescence patterns were complex, the 4D data sets enabled us to track cisternae that could not always be resolved in the 2D projections (Supplementary Fig. S2c). The results were consistent with a cisternal maturation mechanism. Cisternae exhibited green fluorescence for 2–3 min, then passed through a transition phase when both colours were visible, then exhibited red fluorescence for 2–3 min, and finally lost detectable fluorescence (Fig. 3a). Two representative cisternae are shown in Fig. 3 and Supplementary Movie 3a, b. Red-to-green transitions were never observed. The transition events were often abrupt—for example, the cisternae shown in Fig. 3 lost most of their green fluorescence within 15–20 s. As Vrg4 is present in very early Golgi cisternae¹⁹, whereas Sec7 is present in very late Golgi cisternae¹³, the transition phase in which a cisterna contained both fluorescent proteins was brief (Fig. 3b). Each cisterna lost GFP–Vrg4 while acquiring Sec7–DsRed, so the green and red fluorescence signals during the transition phase were low, but both signals were usually visible (Fig. 3c). In four 10-min movies of different cells, we observed a total of 80 green-to-red transitions,

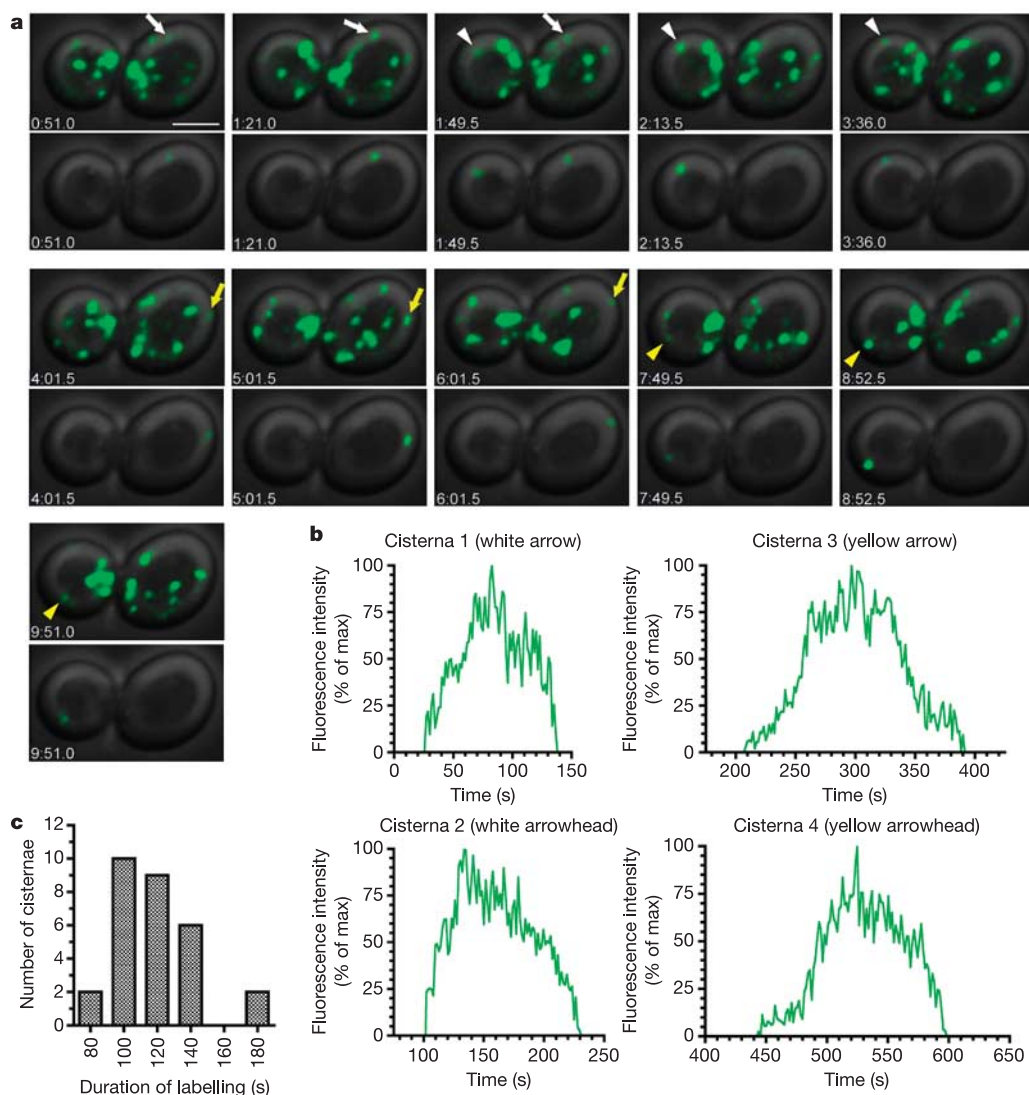


Figure 1 | Sec7-GFP labels each Golgi cisterna for approximately 2 min. **a**, Cropped frames of cells expressing Sec7-GFP from Supplementary Movie 1a, b. Numbers indicate the time (in minutes:seconds) after the movie was initiated. Arrows and arrowheads in the top panels mark four Golgi cisternae that acquired Sec7-GFP fluorescence, increased in brightness and then progressively lost fluorescence. Bottom panels show the same images edited to display only the marked cisternae. Scale bar, 2 μm. **b**, Quantification of

data from Supplementary Movie 1b for the cisternae marked in **a**. The horizontal axis indicates the time after the movie was initiated, and the vertical axis indicates Sec7-GFP fluorescence as a percentage of the maximum signal obtained for the designated cisterna. **c**, Histogram showing the distribution of the duration of labelling of cisternae with Sec7-GFP. The durations of detectable fluorescence signals were measured for 29 cisternae from ten independent movies (see Supplementary Table S1).

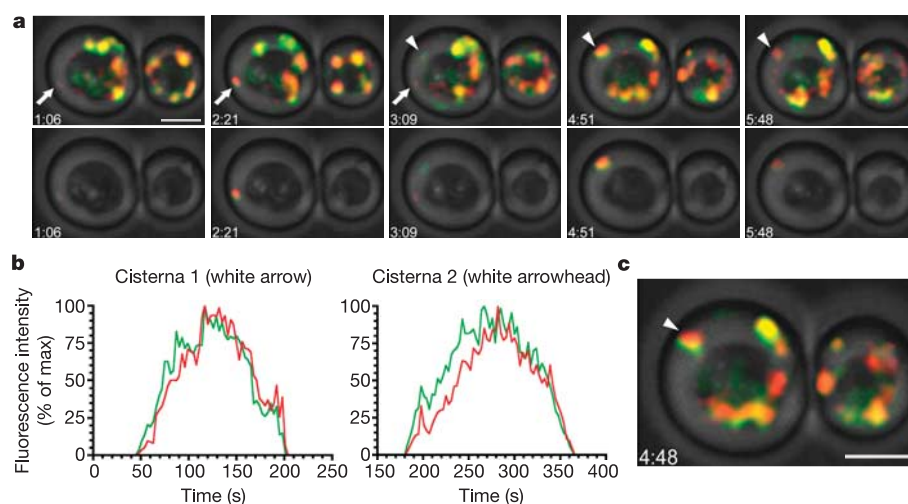


Figure 2 | Two markers of the late Golgi show very similar dynamics. **a**, Cropped frames of cells expressing Sys1-GFP and Sec7-DsRed from Supplementary Movie 2a. **b**, Data are represented as in Fig. 1a. The arrow and arrowhead mark two Golgi cisternae that acquired Sys1-GFP and Sec7-DsRed fluorescence, transiently displayed fluorescence from both markers and then lost fluorescence from both markers. Scale bar, 2 μm.

b, Quantification of data from Supplementary Movie 2b. Data are graphed as in Fig. 1b, except that the vertical axis indicates the fluorescence of Sys1-GFP (green) or Sec7-DsRed (red). **c**, Image from Supplementary Movie 2a showing the apparent partial segregation of Sys1-GFP and Sec7-DsRed within the same cisterna. Scale bar, 2 μm.

indicating that most or all cisternae undergo maturation (Supplementary Table S2). Similar results were seen using Sec7-DsRed in conjunction with GFP-Gos1 or Rbd2-GFP, except that the transition phases were longer than with GFP-Vrg4 (data not shown). During all of these transitions, each cisterna seemed to behave autonomously. In summary, each Golgi cisterna showed sequential labelling with GFP-Vrg4 and Sec7-DsRed, for a total of about 4–6 min. The duration of labelling with these two markers presumably represents most or all of the lifetime of a cisterna.

Do these maturation kinetics match the rate of secretory traffic? It has been proposed that cisternal maturation might be too slow to account for the secretion of certain cargo proteins². To address this issue, we performed radioactive pulse-chase measurements with α -factor, which is one of the most rapidly secreted proteins in

yeast²³. Pro- α -factor exists as a 'core' glycoprotein in the endoplasmic reticulum (ER), and is further glycosylated upon reaching the Golgi, where it is cleaved to generate mature α -factor. To resolve ER export from later transport steps, we used a brief 1-min pulse of ³⁵S-Met to create a population of labelled molecules that had entered the secretory pathway at about the same time. A significant fraction of the label was actually incorporated after the chase began, so the starting time point was defined as 30 s into chase. Pro- α -factor reached the Golgi rapidly, with a half-time of ~1 min (Fig. 4a and Supplementary Fig. S6a). Mature α -factor appeared in the extracellular medium with a half-time of ~7 min. Thus, transport from the early Golgi to the extracellular space required ~6 min. If transport of α -factor through the Golgi occurred by cisternal maturation, our videomicroscopy results suggest that intra-Golgi transport

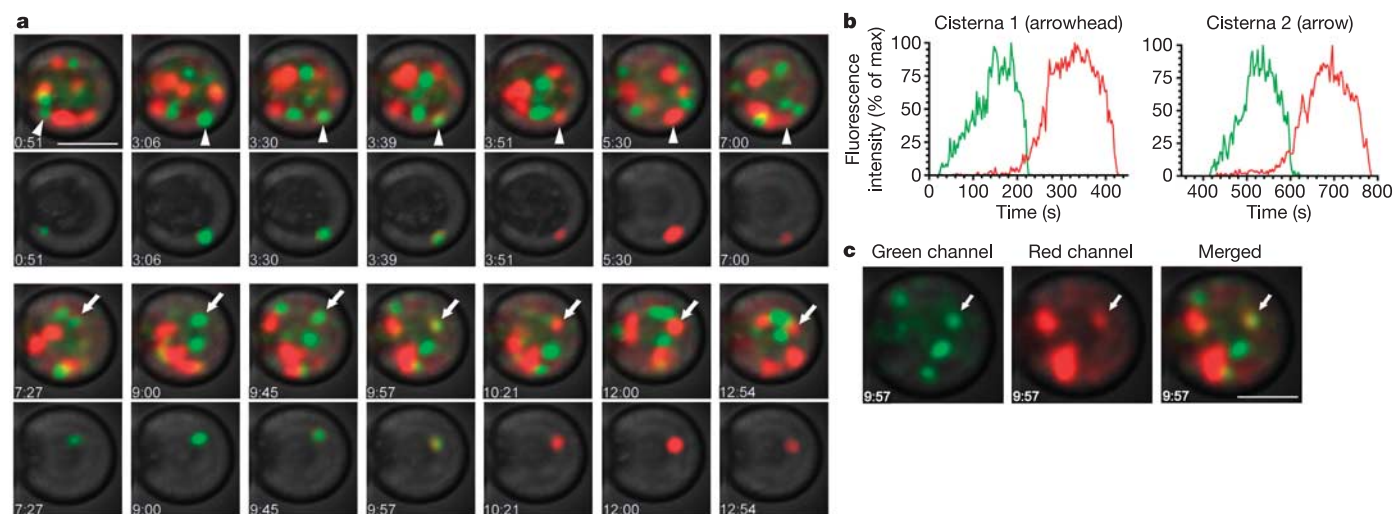


Figure 3 | The resident Golgi protein composition of each cisterna changes over time. **a**, Cropped frames of cells expressing GFP-Vrg4 and Sec7-DsRed from Supplementary Movie 3a. **b**, Data are represented as in Fig. 1a. The arrow and arrowhead mark two Golgi cisternae that acquired GFP-Vrg4 fluorescence, transiently displayed both GFP-Vrg4 and Sec7-DsRed fluorescence, then displayed exclusively Sec7-DsRed

fluorescence and finally lost fluorescence. Scale bar, 2 μm. **b**, Quantification of data from Supplementary Movie 3b. Data are graphed as in Fig. 1b, except that the vertical axis indicates the fluorescence of GFP-Vrg4 (green) or Sec7-DsRed (red). **c**, Example of a cisterna in transition. Green and red channel images for time point 9:57 are shown separately and merged. Scale bar, 2 μm.

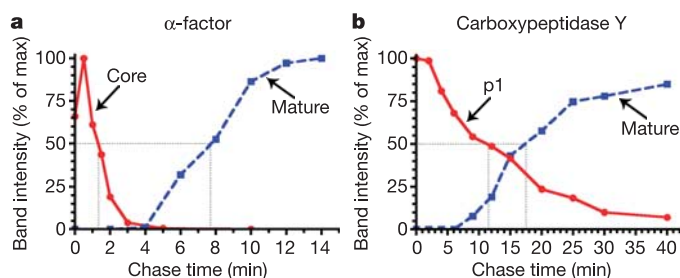


Figure 4 | Secretory pathway transport kinetics are consistent with cisternal maturation. Pulse-chase analysis of α -factor and CPY transport under the growth conditions used for videomicroscopy. See Supplementary Fig. S6 for details. The maximum observed signal for a given protein species was defined as 100%. Dotted lines mark half-times for the appearance and disappearance of the various protein species. **a**, Quantification of α -factor transport. The solid red line represents ER-localized core pro- α -factor, and the dashed blue line represents mature α -factor in the extracellular medium. **b**, Quantification of CPY transport. The solid red line represents ER-localized p1 CPY, and the dashed blue line represents vacuole-localized mature CPY. After 40 min of chase, ~15% of the CPY was still immature, so the maximum value for mature CPY was set at 85%.

required about 4–5 min and Golgi-to-cell-surface transport required an additional 1–2 min. These estimates are consistent with an earlier analysis indicating that pro- α -factor required ~9 min to traverse the Golgi at 15 °C (ref. 23), and with a mammalian cell study indicating that Golgi-to-cell-surface transport is several times faster than intra-Golgi transport⁴. Therefore, our observed rate of cisternal maturation can account for the kinetics of α -factor transport.

We performed similar experiments with carboxypeptidase Y (CPY), which is converted from an ER-localized p1 form to a Golgi-localized p2 form and then to a vacuolar mature form²⁴. Because CPY requires some time to fold in the ER, this kinetic analysis was intrinsically less accurate than with α -factor, but was still feasible using a 2-min pulse. CPY moved from the ER to the Golgi with a half-time of ~11.5 min, and from the ER to the vacuole with a half-time of ~17.5 min, implying that transport from the early Golgi to the vacuole required ~6 min (Fig. 4b and Supplementary Fig. S6b). Using the same logic as above, if transport of CPY through the Golgi occurred by cisternal maturation, then intra-Golgi transport required about 4–5 min and Golgi-to-vacuole transport required an additional 1–2 min. In summary, cisternal maturation was somewhat faster than the transport of α -factor or CPY from the early Golgi to the final post-Golgi destination, indicating that cisternal maturation can account for the transport kinetics of these secretory cargo proteins.

Our data, together with findings of a similar study by another group²⁵, provide direct evidence for the maturation of Golgi cisternae. This conclusion is consistent with earlier analyses of Golgi structure and dynamics in budding yeast^{5,13,26–28}. Future experiments will attempt to visualize secretory cargo proteins simultaneously with Golgi-resident proteins, thereby testing the prediction that secretory cargo proteins remain associated with maturing cisternae. A more daunting challenge will be to elucidate the mechanisms of cisternal maturation. Every Golgi-resident protein presumably recycles from older to younger cisternae, but the recycling pathways are likely to vary. Peripheral membrane proteins such as Sec7 might recycle through the cytosol¹⁰. Some transmembrane Golgi proteins, including certain SNAREs, seem to recycle in COPI vesicles, but there are conflicting reports about whether Golgi glycosylation enzymes enter these vesicles³. Studies of mammalian cells have suggested that proteins might move between cisternae via tubular connections^{29,30}. Although we have not seen evidence for such connections in *S. cerevisiae*, videomicroscopy is presently unable to reveal whether transmembrane Golgi proteins recycle through small vesicles,

larger dissociative transport carriers or inter-cisternal membrane connections. Despite these and many other unresolved questions, the available evidence suggests that in a variety of eukaryotes, cisternal maturation is an important pathway for intra-Golgi transport.

METHODS

Descriptions of yeast strains and plasmids, immunofluorescence and immunoblotting are provided in the Supplementary Information.

Videomicroscopy. *S. cerevisiae* strains were grown to mid-log phase at 23 °C in SD medium, and were imaged under the same conditions. Cells were immobilized on ΔT culture dishes (Bioptechs) using concanavalin A (ref. 15). Single- or dual-colour 4D data sets were collected using a modified LSM 510 confocal microscope (Zeiss), with separate excitation and capture of the red and green fluorescence signals^{15,16}. The pixel size was 0.09 μ m. For each time point, we collected a stack of optical sections spaced 0.375 μ m apart. Typically, 16–24 optical sections were sufficient to span an entire cell. Transmitted light images were stored in the blue channel of RGB TIFF files, and were later converted to grayscale using Adobe Photoshop. Complete z-stacks were collected at time points spaced 1.5 s apart (for single-colour imaging) or 3 s apart (for dual-colour imaging). To reduce photodamage, laser illumination was minimized and confocal scans were carried out as quickly as possible.

The 4D data sets were processed essentially as described¹⁵, primarily using custom macros written with NIH Image 1.63 (<http://rsb.info.nih.gov/ni-image/>). In brief, each optical section was filtered five times with a 3×3 hybrid median filter to reduce shot noise. The resulting z-stacks were then sharpened by iterative deconvolution using Huygens Essential software (Scientific Volume Imaging). Average intensity projections were corrected for photobleaching. Individual cisternae were tracked manually as follows. At any given time during the analysis, a single z-stack of processed optical sections was displayed on the computer screen (see Supplementary Fig. S2). Cisternae of interest were visualized at successive time points by displaying the corresponding z-stacks. A cisterna was deemed suitable for quantification if it could be identified in each of a series of consecutive z-stacks (as in Supplementary Movie S2), and if it was clearly distinct from the other cisternae at every time point. To make edited movies, optical sections containing the cisternae of interest were selected for projection. Fluorescence signals outside the cisternae of interest were erased using a custom macro. This editing was done either in the 2D projections or in individual optical sections. In the figures, fluorescence signals from the movie frames and screenshots have been slightly enhanced using the 'Levels' command in Photoshop.

Pulse-chase analysis. Pulse-chase experiments were performed essentially as described³¹. Briefly, cells grown at 23 °C in SD medium were transferred to SD medium lacking methionine, and were pulsed with 125 μ Ci ml⁻¹ of Trans³⁵S label (MP Biomedicals). Chases were terminated by adding trichloroacetic acid to a final concentration of 5%. Immunoprecipitations were done using a polyclonal anti- α -factor antibody (a gift from T. Graham) or an anti-CPY antibody (a gift from T. Stevens). Samples were analysed by SDS-PAGE using precast gels (Jule, 16% Tris-Tricine gels for α -factor or 4–16% Tris-glycine gradient gels for CPY). Radioactive bands were visualized using a Phosphor-Imager and quantified using ImageQuant software (Molecular Dynamics).

Received 20 December 2005; accepted 7 March 2006.

Published online 14 May 2006; corrected 22 June 2006 (details online).

- Glick, B. S. & Malhotra, V. The curious status of the Golgi apparatus. *Cell* **95**, 883–889 (1998).
- Pelham, H. R. & Rothman, J. E. The debate about transport in the Golgi—two sides of the same coin? *Cell* **102**, 713–719 (2000).
- Rabouille, C. & Klumperman, J. Opinion: The maturing role of COPI vesicles in intra-Golgi transport. *Nature Rev. Mol. Cell Biol.* **6**, 812–817 (2005).
- Bonfanti, L. et al. Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation. *Cell* **95**, 993–1003 (1998).
- Wooding, S. & Pelham, H. R. B. The dynamics of Golgi protein traffic visualized in living yeast cells. *Mol. Biol. Cell* **9**, 2667–2680 (1998).
- Volchuk, A. et al. Megavesicles implicated in the rapid transport of intracisternal aggregates across the Golgi stack. *Cell* **102**, 335–348 (2000).
- Mironov, A. A. et al. Small cargo proteins and large aggregates can traverse the Golgi by a common mechanism without leaving the lumen of cisternae. *J. Cell Biol.* **155**, 1225–1238 (2001).
- Martínez-Menárguez, J. A. et al. Peri-Golgi vesicles contain retrograde but not anterograde proteins consistent with the cisternal progression model of intra-Golgi transport. *J. Cell Biol.* **155**, 1213–1224 (2001).
- Cosson, P., Amherdt, M., Rothman, J. E. & Orci, L. A resident Golgi protein is excluded from peri-Golgi vesicles in NRK cells. *Proc. Natl Acad. Sci. USA* **99**, 12831–12834 (2002).

10. Franzusoff, A., Redding, K., Crosby, J., Fuller, R. S. & Schekman, R. Localization of components involved in protein transport and processing through the yeast Golgi apparatus. *J. Cell Biol.* **112**, 27–37 (1991).
11. Preuss, D., Mulholland, J., Franzusoff, A., Segev, N. & Botstein, D. Characterization of the *Saccharomyces* Golgi complex through the cell cycle by immunoelectron microscopy. *Mol. Biol. Cell* **3**, 789–803 (1992).
12. Payne, W. E., Gannon, P. M. & Kaiser, C. A. An inducible acid phosphatase from the yeast *Pichia pastoris*: characterization of the gene and its product. *Gene* **163**, 19–26 (1995).
13. Mogelsvang, S., Gomez-Ospina, N., Soderholm, J., Glick, B. S. & Staehelin, L. A. Tomographic evidence for continuous turnover of Golgi cisternae in *Pichia pastoris*. *Mol. Biol. Cell* **14**, 2277–2291 (2003).
14. Duden, R. & Schekman, R. in *The Golgi Apparatus* (eds Berger, E. G. & Roth, J.) 219–246 (Birkhäuser Verlag, Basel, 1997).
15. Bevis, B. J., Hammond, A. T., Reinke, C. A. & Glick, B. S. *De novo* formation of transitional ER sites and Golgi structures in *Pichia pastoris*. *Nature Cell Biol.* **4**, 750–756 (2002).
16. Hammond, A. T. & Glick, B. S. Raising the speed limits for 4D fluorescence microscopy. *Traffic* **1**, 935–940 (2000).
17. Rossanese, O. W. *et al.* A role for actin, Cdc1p and Myo2p in the inheritance of late Golgi elements in *Saccharomyces cerevisiae*. *J. Cell Biol.* **153**, 47–61 (2001).
18. Dean, N., Zhang, Y. B. & Poster, J. B. The *VRG4* gene is required for GFP-mannose transport into the lumen of the golgi in the yeast, *Saccharomyces cerevisiae*. *J. Biol. Chem.* **272**, 31908–31914 (1997).
19. Abe, M., Noda, Y., Adachi, H. & Yoda, K. Localization of GDP-mannose transporter in the Golgi requires retrieval to the endoplasmic reticulum depending on its cytoplasmic tail and coatomer. *J. Cell Sci.* **117**, 5687–5696 (2004).
20. Behnia, R., Panic, B., Whyte, J. R. & Munro, S. Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. *Nature Cell Biol.* **6**, 405–413 (2004).
21. McNew, J. A. *et al.* Gos1p, a *Saccharomyces cerevisiae* SNARE protein involved in Golgi transport. *FEBS Lett.* **435**, 89–95 (1998).
22. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
23. Brigance, W. T., Barlowe, C. & Graham, T. R. Organization of the yeast Golgi complex into at least four functionally distinct compartments. *Mol. Biol. Cell* **11**, 171–182 (2000).
24. Horazdovsky, B. F., DeWald, D. B. & Emr, S. D. Protein transport to the yeast vacuole. *Curr. Opin. Cell Biol.* **7**, 544–551 (1995).
25. Matsuura-Tokita, K., Takeuchi, M., Ichihara, A., Mikuriya, K. & Nakano, A. Live imaging of yeast Golgi cisternal maturation. *Nature* advance online publication, doi:10.1038/nature04737 (14 May 2006).
26. Morin-Ganet, M.-N., Rambourg, A., Deitz, S. B., Franzusoff, A. & Képès, F. Morphogenesis and dynamics of the yeast Golgi apparatus. *Traffic* **1**, 56–68 (2000).
27. Todorow, Z., Spang, A., Carmack, E., Yates, J. & Schekman, R. Active recycling of yeast Golgi mannosyltransferase complexes through the endoplasmic reticulum. *Proc. Natl Acad. Sci. USA* **97**, 13643–13648 (2000).
28. Reinke, C. A., Kozik, P. & Glick, B. S. Golgi inheritance in *Saccharomyces cerevisiae* depends on ER inheritance. *Proc. Natl Acad. Sci. USA* **101**, 18018–18023 (2004).
29. Trucco, A. *et al.* Secretory traffic triggers the formation of tubular continuities across Golgi sub-compartments. *Nature Cell Biol.* **6**, 1071–1081 (2004).
30. Marsh, B. J., Volkmann, N., McIntosh, J. R. & Howell, K. E. Direct continuities between cisternae at different levels of the Golgi complex in glucose-stimulated mouse islet beta cells. *Proc. Natl Acad. Sci. USA* **101**, 5565–5570 (2004).
31. Graham, T. R. in *Current Protocols in Cell Biology* (ed. Morgan, K.) 7.6.1–7.6.9 (Wiley, New York, 2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Thanks to A. Franzusoff, T. Stevens and P. Silver for providing reagents, to A. Hammond for advice about microscopy, and to T. Graham for help with pulse-chase analysis. We are grateful to A. Nakano for discussions and for sharing data before publication. This work was supported by grants from the March of Dimes Birth Defects Foundation, the National Institutes of Health and the American Cancer Society.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.S.G. (bsglick@midway.uchicago.edu).

Live imaging of yeast Golgi cisternal maturation

Kumi Matsuura-Tokita¹, Masaki Takeuchi^{1†}, Akira Ichihara^{1,2}, Kenta Mikuriya² & Akihiko Nakano^{1,3}

There is a debate over how protein trafficking is performed through the Golgi apparatus^{1–4}. In the secretory pathway, secretory proteins that are synthesized in the endoplasmic reticulum enter the early compartment of the Golgi apparatus called *cis* cisternae, undergo various modifications and processing, and then leave for the plasma membrane from the late (*trans*) cisternae. The cargo proteins must traverse the Golgi apparatus in the *cis*-to-*trans* direction. Two typical models propose either vesicular transport or cisternal progression and maturation for this process. The vesicular transport model predicts that Golgi cisternae are distinct stable compartments connected by vesicular traffic, whereas the cisternal maturation model predicts that cisternae are transient structures that form *de novo*, mature from *cis* to *trans*, and then dissipate. Technical progress in live-cell imaging has long been awaited to address this problem. Here we show, by the use of high-speed three-dimensional confocal microscopy, that yeast Golgi cisternae do change the distribution of resident membrane proteins from the *cis* nature to the *trans* over time, as proposed by the maturation model, in a very dynamic way.

We have been developing a high-performance confocal laser scanning microscopic system, by combining the spinning-disk (Nipkow disk) confocal scanning method and HARP (high-gain avalanche rushing amorphous photoconductor) cameras⁵. The most recent prototype features multi-colour observation with extremely high sensitivity and high signal-to-noise ratio, permitting the high-speed observation of weak fluorescent signals within a living cell. We applied this technology to the observation of the Golgi apparatus in yeast cells.

The yeast *Saccharomyces cerevisiae* has a unique feature in the organization of its Golgi. Unlike higher animal and plant cells, in which the Golgi apparatus forms a stacked structure of several cisternae, *S. cerevisiae* usually shows single layers of Golgi cisterna scattering in the cytoplasm^{6–8}. The clustering of the Golgi in the perinuclear region, seen in mammalian cells, does not take place in yeast cells; yeast therefore provides an ideal system in which to view individual Golgi cisternae under a microscope⁹.

Figure 1 shows the yeast cells simultaneously expressing two Golgi markers with different fluorescence colours as revealed by a confocal microscope. In Fig. 1a, Rer1, a retrieval receptor for membrane proteins of the endoplasmic reticulum^{10,11}, was fused to the green fluorescent protein (GFP) and used as a *cis*-Golgi marker. Gos1 is a SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) molecule in a later compartment of yeast Golgi^{12,13}, and was used as a medial marker in the form of a fusion with the monomeric red fluorescent protein (mRFP)¹⁴. In Fig. 1b, Sed5, another SNARE molecule residing mainly in the *cis*-Golgi^{15,16}, was fused to mRFP, and Sec7, a guanine-nucleotide exchange factor for Arf GTPases in the *trans*-Golgi^{17,18}, was fused to GFP. These fusion proteins were selected out of many constructs we had made, because they proved fully functional by complementation of mutant pheno-

types without affecting the morphology of the Golgi apparatus. In either of the two combinations, green and red signals were mostly discrete, confirming that early and late Golgi cisternae scatter separately in the yeast cytoplasm. Effects of traffic mutations on

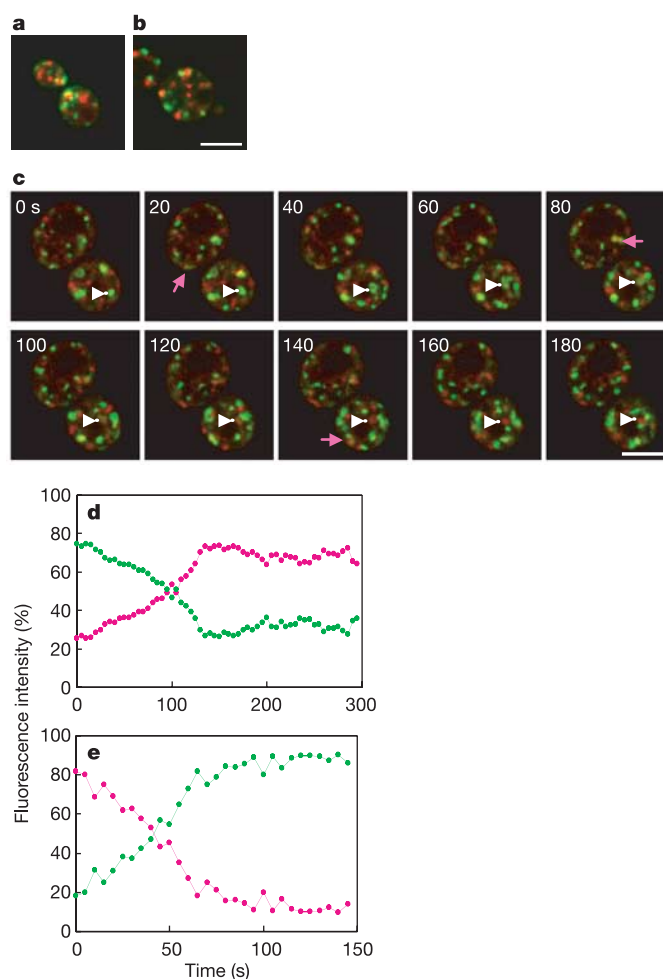


Figure 1 | Dual-fluorescence observation of yeast Golgi cisternae.

a, Wild-type cells expressing GFP-Rer1 (*cis*, green) and mRFP-Gos1 (medial, red). **b**, Wild-type cells expressing mRFP-Sed5 (*cis*, red) and Sec7-GFP (*trans*, green). **c**, Cells as in **a** were subjected to sequential recording at the rate of one frame per 5 s (see Supplementary movie 1). Frames selected at intervals of 20 s are shown. **d**, Analysis of colour change of the cisterna marked by the white arrowheads in **c**. Relative fluorescence intensities in the green and red channels (points coloured accordingly) were normalized to the total and are plotted against time. **e**, Similar analysis of colour change from the experiment using cells as in **b**. Scale bars, 5 μ m.

¹Molecular Membrane Biology Laboratory, RIKEN Discovery Research Institute, Wako, Saitama 351-0198, Japan. ²Biocenter, Yokogawa Electric Corporation, Musashino, Tokyo 180-8750, Japan. ³Department of Biological Sciences, Graduate School of Science, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. [†]Present address: Department of Molecular Structure, Institute for Molecular Science, National Institutes of Natural Sciences, Okazaki, Aichi 444-8585, Japan.

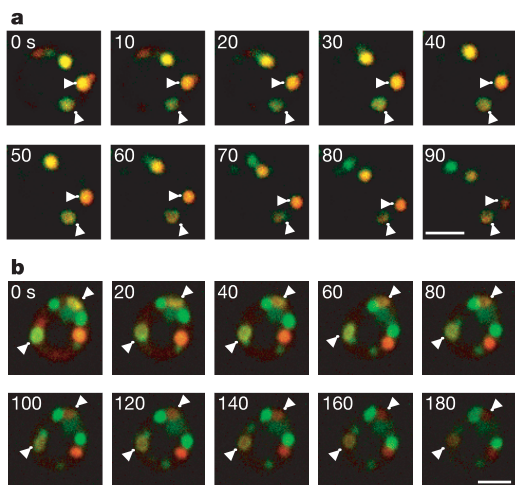


Figure 2 | Dual fluorescence observation with the α COP mutant *ret1-1*. **a**, *ret1-1* cells expressing GFP-Gos1 (medial, green) and Sec7-mRFP (*trans*, red) were preincubated at the restrictive temperature 37 °C for 30 min and then observed under a microscope whose stage was kept at 37 °C. Images were taken at the rate of one frame per 5 s. Frames at intervals of 10 s are shown. White arrowheads indicate the cisternae whose colour changed. **b**, Similar analysis to that in **a**. Frames at intervals of 20 s are shown. Scale bars, 2 μ m. See also Supplementary movies 2 and 3.

the localizations of these markers are shown in Supplementary Fig. S2.

Next we attempted to follow any changes of individual cisternae in real time. If the classic vesicular transport model is correct, *cis*, medial and *trans* cisternae should be stable compartments, and their fluorescence colours marked by resident proteins should stay unchanged. However, what we observed was a clear change of colour over time. Figure 1c shows a typical example of yeast cells expressing GFP-Rer1 (*cis*, green) and mRFP-Gos1 (medial, red) (see also Supplementary movie 1). In this sequence of 3 min, one cisterna (white arrowheads), which was green at the beginning, changed its colour to yellow and then to red. This indicates that the cisterna, which was occupied by GFP-Rer1 at time zero, became dominated by mRFP-Gos1 during this period. Other examples of colour change of cisternae can also be seen in the same sequence (pink arrows). Figure 1d shows the time course of apparent colour change for the cisterna marked by the white arrowheads, by plotting relative fluorescence intensities observed in the green and red channels of the camera. It took about 2 min for the conversion from the *cis* to the medial property. Conversely, a colour change of red to green was observed for the cells expressing mRFP-Sed5 (*cis*, red) and Sec7-GFP (*trans*, green). A result of similar time-course analysis is shown in Fig. 1e. These colour changes were always unidirectional; the reverse change of colour was never observed. One more example of colour change for the pair of markers GFP-Gos1 (medial) and Sec7-mRFP (*trans*) is shown in Supplementary Fig. S3. In this case the frequency of colour switch was higher than that in Fig. 1c, perhaps because the extent of overlap was larger for these two residents.

We next asked how these membrane components are transferred between cisternae. Cisternal maturation models often postulate retrograde vesicular traffic from later cisternae to earlier ones, which carry back Golgi-resident proteins. A candidate of such vesicular carriers is COPI vesicles. We examined whether a mutation of COPI coat affects the colour change of yeast Golgi. We chose the α COP mutant *ret1-1*, which was known to be defective in COPI vesicle formation at the restrictive temperature 37 °C (ref. 19). As shown in Fig. 2, the medial cisternae marked with GFP-Gos1 changed colour to red, representing Sec7-mRFP (*trans*) in *ret1-1*

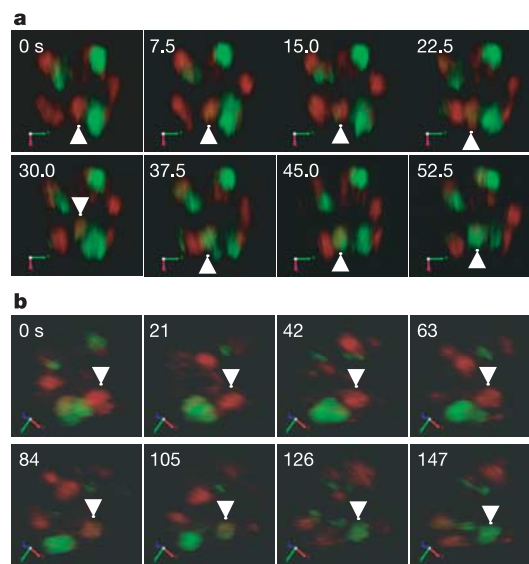


Figure 3 | Three-dimensional observation of yeast Golgi cisternae. Wild-type yeast cells expressing mRFP-Sed5 (*cis*, red) and Sec7-GFP (*trans*, green) were observed under a high-speed confocal microscope with scanning in the *z* axis. **a**, Five optical slices 1 μ m apart were taken at 0.5-s intervals to cover a whole cell. Three-dimensional images were reconstructed into a movie (Supplementary movie 4), of which representative views are shown. **b**, As in **a** except that seven optical slices 1 μ m apart were taken at 1-s intervals (Supplementary movie 5). Scale bars, 2 μ m.

cells at the restrictive temperature (see also Supplementary movies 2 and 3). However, the speed of transition seemed significantly lower in the *ret1-1* mutant. To obtain a rough estimate of the transition rates, we measured the red/green fluorescence ratio during the colour conversion and calculated its doubling time. The colour change of cisternae in wild-type cells at 37 °C took a ratio doubling time of 27 ± 7 s ($\pm$ s.d.; $n = 11$), whereas in *ret1-1* cells at 37 °C it took 92 ± 39 s ($n = 14$). These observations indicate that COPI vesicles are important in the conversion of cisternae but that other mechanism(s) may also operate.

In the settings of our spinning-disk confocal microscope, the thickness of the confocal plane was 1–2 μ m; this did not cover the whole yeast cell, whose diameter was about 5 μ m. This problem could be resolved by moving the *z* axis of the microscope, collecting multiple confocal images and reconstructing the three-dimensional (3D) architecture. The sequences of pictures in Fig. 3 show whole 3D images of yeast cells expressing mRFP-Sed5 (*cis*, red) and Sec7-GFP (*trans*, green) (Supplementary movies 4 and 5). If necessary, the projected images were appropriately tilted to minimize the overlapping of individual cisternae. Real-time 3D observation of the Golgi cisternae again revealed dynamic features of their conversion. Single isolated cisternae changed their properties over time (white arrowheads).

We also tried to improve the spatial resolution of confocal images. The deconvolution technique was quite powerful in making 3D-reconstructed observations. Two typical results on the cells expressing medial (mRFP-Gos1) and *trans* (Sec7-GFP) markers are shown in Fig. 4 and Supplementary movies 6 and 7. With deconvolution, Golgi cisternae no longer look like flat discs but display complex structures. An enlargement of the zero-time picture of Fig. 4a is shown in Fig. 4b. The Golgi cisternae look mostly fenestrated and are sometimes reminiscent of the tubular network proposed by studies by electron microscopy^{7,20,21}. With this improved resolution, the conversion of colours from red to green was even more marked. In the intermediate state of transition there seemed to be a segregation of domains. In the example shown in Fig. 4c, almost all the cisternae observed at the beginning exhibited

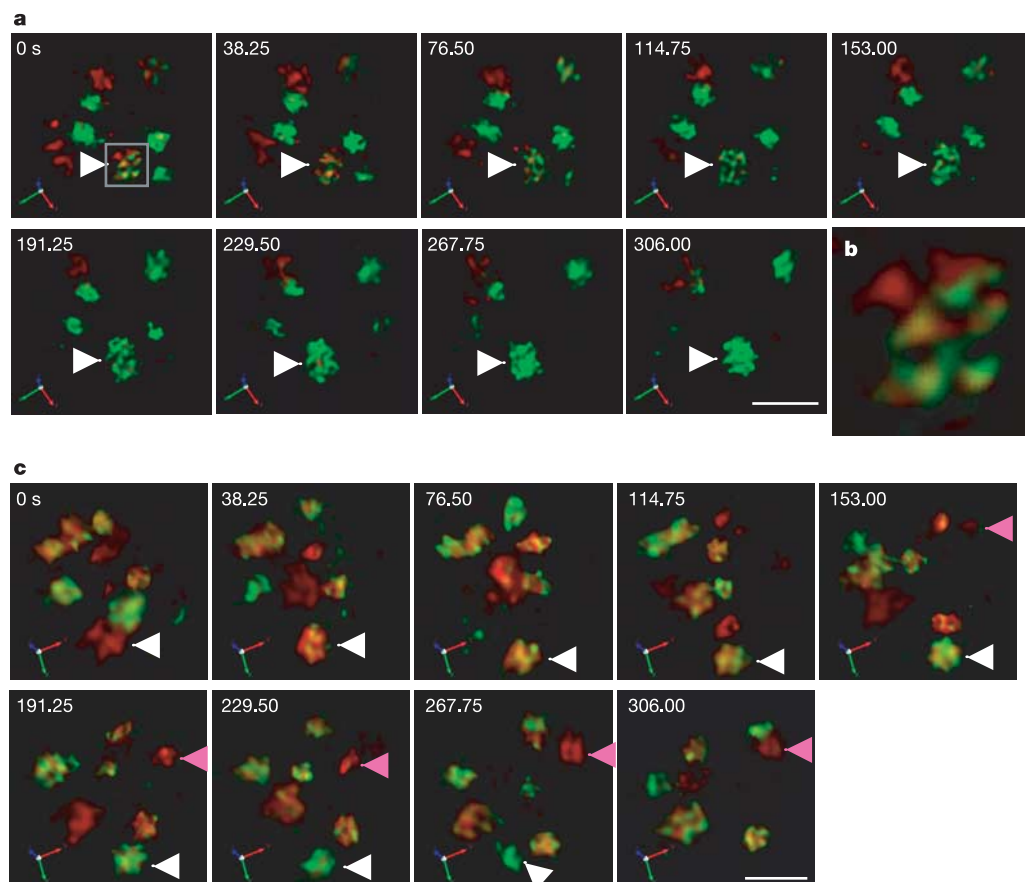


Figure 4 | Three-dimensional deconvolution observation of yeast Golgi cisternae. Wild-type yeast cells expressing mRFP–Gos1 (medial, red) and Sec7–GFP (*trans*, green) were observed under a high-speed confocal microscope. **a**, Fifty optical slices 0.1 μm apart were taken at 0.25-s intervals to cover a whole cell. Three-dimensional images were reconstructed, deconvolved by the parameters optimized for the spinning-disk confocal

scanner and organized into a movie (Supplementary movie 6). Representative views of the movies are shown here. **b**, An enlargement of the zero-time picture of sequence **a** (indicated by the white square). **c**, As in **a** except that 50 optical slices 0.14 μm apart were taken (Supplementary movie 7). Scale bars, 2 μm .

the colour conversion within 5 min. The disappearance of green cisternae (white arrowheads) and the new appearance of red cisternae (pink arrowheads) are also clearly visible.

Our data showing that the αCOP mutant slows down but does not completely block the colour conversion of cisternae (Fig. 2) indicate not only important roles for COPI vesicles but also the presence of other mechanisms. For example, tubular connection has recently been shown to form between Golgi cisternae in mammalian cells^{22,23}. Indeed, in our experiments (Figs 3 and 4) we often observed the close approach of two distinct cisternae to each other, indicating a possible direct contact between them. Improvement of our imaging methodology should be further pursued to provide an unambiguous understanding of how membrane components are conveyed between cisternae. Quantitative kinetic analysis with a variety of mutants will help us dissect the contributions of different mechanisms.

Our observations, together with a similar study from Glick's group²⁴, have clearly indicated that Golgi cisternae do mature over time in yeast cells (see Supplementary Fig. S1 for our model). The contribution to the sorting events of membrane segregation during maturation will certainly be an issue to be pursued next. Many more questions remain to be answered. The rate of maturation seems to be fast enough to explain the overall transit of secretory cargo. However, given that the rates of transport may vary from cargo to cargo, it will be necessary to examine how various cargo proteins are travelling in maturing cisternae. The mechanism of unidirectional maturation also needs to be explained. The technology of live-cell imaging is

still in progress. The challenge of resolution limits in time and space will certainly be a way of approaching these further exciting problems.

METHODS

Two microscopic settings were used for confocal fluorescence imaging of living cells. For normal two-dimensional observations, an Olympus BX-52 microscope was equipped with a CSU10 spinning-disk confocal scanner (Yokogawa Electric Corporation) and an ORCA-3CCD charge-coupled device colour camera (Hamamatsu Photonics). Images were analysed with AquaCosmos software (Hamamatsu Photonics). For higher-speed observations, a custom-made system produced by the Dynamic-Bio Project was used. In this system, an Olympus IX-70 microscope was equipped with a special high-signal-to-noise-ratio colour confocal system (Yokogawa Electric), image intensifiers (Hamamatsu Photonics) and high-speed and ultra-high-sensitivity HARP cameras (NHK Engineering Service and Hitachi Kokusai Electric). In both settings, an Ar^+/Kr^+ laser (Melles Griot) was used to excite GFP at 488 nm and mRFP at 568 nm simultaneously. For 3D imaging, the objective lens was oscillated vertically to the sample plane by a piezo actuator system (Yokogawa Electric). Collected pictures were analysed with Volocity software (Improvision). Deconvolution analysis was also performed with Volocity by using theoretical point-spread functions optimized for CSU10.

To test the effect of a COPI mutation, GFP–Gos1 and Sec7–mRFP were expressed in wild-type and *ret1-1* cells. Cells grown to exponential phase at a permissive temperature (23 °C) were mounted on a glass slide and kept at a restrictive temperature (37 °C) on the microscope stage by the use of a thermo-control system (Tokai Hit Co.). Images were acquired 30–60 min after the temperature shift. Fluorescence intensities of the cisternae that were undergoing a green-to-red colour change during observation were quantified with AquaCosmos.

Eleven cisternae from 8 wild-type cells and 14 cisternae from 7 *ret1-1* cells were subjected to kinetic analysis.

Received 14 December 2005; accepted 22 March 2006.

Published online 14 May 2006.

1. Glick, B. S. & Malhotra, V. The curious status of the Golgi apparatus. *Cell* **95**, 883–889 (1998).
2. Pelham, H. R. B. & Rothman, J. E. The debate about transport in the Golgi—two sides of the same coin? *Cell* **102**, 713–719 (2000).
3. Pelham, H. Getting stuck in the Golgi. *Traffic* **1**, 191–192 (2000).
4. Pelham, H. R. B. Traffic through the Golgi apparatus. *J. Cell Biol.* **155**, 1099–1101 (2001).
5. Nakano, A. Spinning-disk confocal microscopy—a cutting-edge tool for imaging of membrane traffic. *Cell Struct. Funct.* **27**, 349–355 (2002).
6. Preuss, D. *et al.* Characterization of the *Saccharomyces* Golgi complex through the cell cycle by immunoelectron microscopy. *Mol. Biol. Cell* **3**, 789–803 (1992).
7. Rambourg, A., Clermont, Y. & Képès, F. Modulation of the Golgi apparatus in *Saccharomyces cerevisiae* *sec7* mutants as seen by three-dimensional electron microscopy. *Anat. Rec.* **237**, 441–452 (1993).
8. Rossanese, O. W. *et al.* Golgi structure correlates with transitional endoplasmic reticulum organization in *Pichia pastoris* and *Saccharomyces cerevisiae*. *J. Cell Biol.* **145**, 69–81 (1999).
9. Nakano, A. Yeast Golgi apparatus—dynamics and sorting. *Cell. Mol. Life Sci.* **61**, 186–192 (2004).
10. Sato, K., Sato, M. & Nakano, A. Rer1p as common machinery for the endoplasmic reticulum localization of membrane proteins. *Proc. Natl Acad. Sci. USA* **94**, 9693–9698 (1997).
11. Sato, K., Sato, M. & Nakano, A. Rer1p, a retrieval receptor for endoplasmic reticulum membrane proteins, is dynamically localized to the Golgi apparatus by coatamer. *J. Cell Biol.* **152**, 935–944 (2001).
12. McNew, J. A. *et al.* Gos1p, a *Saccharomyces cerevisiae* SNARE protein involved in Golgi transport. *FEBS Lett.* **435**, 89–95 (1998).
13. Bevis, B. J., Hammond, A. T., Reinke, C. A. & Glick, B. S. *De novo* formation of transitional ER sites and Golgi structures in *Pichia pastoris*. *Nature Cell Biol.* **4**, 750–756 (2002).
14. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 7877–7882 (2002).
15. Hardwick, K. G. & Pelham, H. R. *SED5* encodes a 39-kD integral membrane protein required for vesicular transport between the ER and the Golgi complex. *J. Cell Biol.* **119**, 513–521 (1992).
16. Wooding, S. & Pelham, H. R. B. The dynamics of Golgi protein traffic visualized in living yeast cells. *Mol. Biol. Cell* **9**, 2667–2680 (1998).
17. Achstetter, T., Franzusoff, A., Field, C. & Schekman, R. SEC7 encodes an unusual, high molecular weight protein required for membrane traffic from the yeast Golgi apparatus. *J. Biol. Chem.* **263**, 11711–11717 (1988).
18. Sata, M., Donaldson, J. G., Moss, J. & Vaughan, M. Brefeldin A-inhibited guanine nucleotide-exchange activity of Sec7 domain from yeast Sec7 with yeast and mammalian ADP ribosylation factors. *Proc. Natl Acad. Sci. USA* **95**, 4204–4208 (1998).
19. Letourneur, F. *et al.* Coatamer is essential for retrieval of dilysine-tagged proteins to the endoplasmic reticulum. *Cell* **79**, 1199–1207 (1994).
20. Rambourg, A., Clermont, Y., Ovtracht, L. & Kepes, F. Three-dimensional structure of tubular networks, presumably Golgi in nature, in various yeast strains: a comparative study. *Anat. Rec.* **243**, 283–293 (1995).
21. Rambourg, A., Jackson, C. L. & Clermont, Y. Three dimensional configuration of the secretory pathway and segregation of secretion granules in the yeast *Saccharomyces cerevisiae*. *J. Cell Sci.* **114**, 2231–2239 (2001).
22. Marsh, B. J., Volkmann, N., McIntosh, J. R. & Howell, K. E. Direct continuities between cisternae at different levels of the Golgi complex in glucose-stimulated mouse islet beta cells. *Proc. Natl Acad. Sci. USA* **101**, 5565–5570 (2004).
23. Trucco, A. *et al.* Secretory traffic triggers the formation of tubular continuities across Golgi sub-compartments. *Nature Cell Biol.* **6**, 1071–1081 (2004).
24. Losev, E. *et al.* Golgi maturation visualized in living yeast. *Nature* advance online publication, doi:10.1038/nature04717 (14 May 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all the members of the Nakano group of the Dynamic-Bio Project for the accomplishment of this microscopy project; Yokogawa Electric Corporation, NHK (Japan Broadcasting Corporation), NHK Engineering Service, Hitachi Kokusai Electric, and the Research Association for Biotechnology for their contributions; R. Tsien for the distribution of mRFP; Olympus Corporation for technical help; and B. Glick for exchange of information before publication. This work was supported by national funds from the Ministry of Economy, Trade and Industry of Japan and the New Energy and Industrial Technology Development Organization, partly by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and partly by the funds from the Bioarchitect, the Real-Time Bionanomachine and the Extreme Photonics Projects of RIKEN.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.N. (nakano@riken.jp).

Increased cell-to-cell variation in gene expression in ageing mouse heart

Rumana Bahar¹, Claudia H. Hartmann², Karl A. Rodriguez³, Ashley D. Denny³, Rita A. Busuttil¹, Martijn E. T. Dollé⁴, R. Brent Calder¹, Gary B. Chisholm³, Brad H. Pollock³, Christoph A. Klein² & Jan Vijg¹

The accumulation of somatic DNA damage has been implicated as a cause of ageing in metazoa^{1,2}. One possible mechanism by which increased DNA damage could lead to cellular degeneration and death is by stochastic deregulation of gene expression. Here we directly test for increased transcriptional noise in aged tissue by dissociating single cardiomyocytes from fresh heart samples of both young and old mice, followed by global mRNA amplification and quantification of mRNA levels in a panel of housekeeping and heart-specific genes. Although gene expression levels already varied among cardiomyocytes from young heart, this heterogeneity was significantly elevated at old age. We had demonstrated previously an increased load of genome rearrangements and other mutations in the heart of aged mice^{3,4}. To confirm that increased stochasticity of gene expression could be a result of increased genome damage, we treated mouse embryonic fibroblasts in culture with hydrogen peroxide. Such treatment resulted in a significant increase in cell-to-cell variation in gene expression, which was found to parallel the induction and persistence of genome rearrangement mutations at a *lacZ* reporter locus. These results underscore the stochastic nature of the ageing process, and could provide a mechanism for age-related cellular degeneration and death in tissues of multicellular organisms.

With the emergence of *in situ* techniques it has become possible to visualize mRNA levels in single cells. *In situ* methods have even allowed the visualization of specific sites of transcription of multiple genes simultaneously⁵. However, accurate quantification and the demonstration of significant differences in the expression levels of different mRNAs in cells freshly derived from mammalian tissues have, thus far, not been widely attempted⁶. Here we use a modification of a highly accurate, unbiased method⁷ to amplify global mRNA from enzymatically dissociated cardiomyocytes from young and old mouse heart and quantify transcript levels of arbitrarily chosen genes by real-time polymerase chain reaction (PCR). The genes selected were seven housekeeping genes, three heart-specific genes, two protease-encoding genes, and three mitochondrial genes (Table 1). Because any gene had the potential to undergo random changes in its transcript level, we used different nuclear and mitochondrial genes as reference genes, reasoning that although this would not allow us to draw conclusions about absolute differences, our basic hypothesis of a significant age-related increase in cell-to-cell variation in gene expression remained perfectly testable.

Heart cardiomyocytes were isolated from a similar area of ventricular heart tissue of young (6 months) and old (27 months) male mice. Dissociated cells contained over 90% rod-shaped, striated myocytes, which were easily recognizable under the microscope as cardiomyocytes (Fig. 1a). There was no evidence of the typical signs of either apoptotic or necrotic cell death—that is, rounded-up cells with plasma membrane blebs (apoptosis), or cell membrane disruption,

extensive vacuolization or swollen nuclei (necrosis)⁸. Viability of virtually all cells from both young and old mice was also indicated by Trypan-blue dye exclusion.

To ensure that the expression of our target genes was stable during the isolation procedure, we compared heart tissue that was flash-frozen immediately after resection with a similar heart sample that was subjected to the isolation procedure—that is, sliced with a razor

Table 1 | Significance of age-related increase in gene expression variation

Genes	Significance	Normalizers
Housekeeping		
<i>Gapdh</i>	$P < 0.0001$	<i>Actb</i>
	$P < 0.0001$	<i>B2m</i>
<i>Actb</i>	$P < 0.0001$	<i>Gapdh</i>
	$P < 0.0001$	<i>B2m</i>
<i>B2m</i>	$P < 0.0001$	<i>Gapdh</i>
	$P < 0.0001$	<i>Actb</i>
<i>Tuba6</i>	$P < 0.0001$	<i>Gapdh</i>
	$P < 0.0001$	<i>Actb</i>
	$P < 0.0001$	<i>B2m</i>
<i>Lmna</i>	$P = 0.010$	<i>Gapdh</i>
	$P = 0.1131$	<i>Actb</i>
	$P = 0.1027$	<i>B2m</i>
<i>Cox6a2</i>	$P < 0.0001$	<i>Cox7a1</i>
	$P < 0.0001$	<i>Myl2</i>
<i>Cox7a1</i>	$P < 0.0001$	<i>Cox6a2</i>
	$P < 0.0001$	<i>Myl2</i>
Heart-specific		
<i>Actc1</i>	$P < 0.0001$	<i>Gapdh</i>
	$P < 0.0001$	<i>Actb</i>
	$P < 0.0001$	<i>B2m</i>
<i>Lpl</i>	$P < 0.0001$	<i>Gapdh</i>
	$P < 0.0001$	<i>Actb</i>
	$P < 0.0001$	<i>B2m</i>
<i>Myl2</i>	$P < 0.0001$	<i>Cox6a2</i>
	$P < 0.0001$	<i>Cox7a1</i>
Proteases		
<i>Cst3</i>	$P < 0.0001$	<i>Gapdh</i>
	$P < 0.0001$	<i>Actb</i>
	$P < 0.0001$	<i>B2m</i>
<i>Ctss</i>	$P = 0.017$	<i>Gapdh</i>
	$P = 0.7761$	<i>Actb</i>
	$P = 0.8073$	<i>B2m</i>
Mitochondrial		
<i>COX1</i>	NS	<i>COX2</i>
	NS	<i>COX3</i>
<i>COX2</i>	NS	<i>COX1</i>
	NS	<i>COX3</i>
<i>COX3</i>	NS	<i>COX1</i>
	NS	<i>COX2</i>

The significance of increased cell-to-cell variation in gene expression in young versus old mouse heart was determined using Bartlett's test. NS, not significant.

¹Buck Institute for Age Research, Novato, California 94945, USA. ²Institut für Immunologie, Ludwig-Maximilians Universität, 80336 München, Germany. ³University of Texas Health Science Center, San Antonio, Texas 78245, USA. ⁴National Institute of Public Health and the Environment, 3720 BA Bilthoven, The Netherlands.

blade, incubated for 30 min with collagenase at room temperature ($\sim 20^{\circ}\text{C}$) and then flash-frozen. A small amount of each tissue was amplified according to the same protocol. There was no significant effect of either the enzymatic treatment or the age of the animal on the expression levels of the target genes (Supplementary Table 1).

To verify the reproducibility of the amplification method in maintaining correct representation of individual mRNA levels, we repeatedly amplified the equivalent of a single cell from a pool of lysed cardiomyocytes, as compared with amplifications of single cells, from the heart of a 6-month-old mouse (Fig. 1b). We subsequently quantified for each cell, or single-cell equivalent, the expression level of the heart-specific gene *Myl2*, encoding the myosin ventricular regulatory light chain. In this case, the highly expressed mitochondrial cytochrome *c* oxidase gene *COX1* (also known as *mt-Co1*) was used as the reference gene. The results show very little variation in relative quantity among single-cell equivalents (Fig. 1c), which was within the range of the real-time PCR error. In contrast, variation in the relative quantity of *Myl2* among single cells was significant, indicating cell-to-cell variation in gene expression already occurs among cells from a young heart. Similar results were obtained with other highly expressed genes. Notably, transcriptional noise among single-cell equivalents was significant when studying genes expressed at a lower level in the heart, such as *Actb* or *Gapdh*, probably because of increased sensitivity of transcripts to stochastic fluctuations when present at low copy number⁹.

We then captured at least 15 individual cardiomyocytes for each of three young and three old animals, and subjected each cell to the global mRNA amplification procedure followed by quantification of the expression level of each target gene by real-time PCR. For normalization, we used different genes in the panel depending on the expression level (Table 1). Bartlett's test for unequal variances was used to test for differences in gene expression variability. The results indicated a highly significant increase in cell-to-cell variation in the expression levels of all nuclear genes in the old animals as compared with the young ones (Table 1). This increased cell-to-cell variation was observed irrespective of the reference gene used; however, for the genes *Lmna* and *Ctss* significance levels were reached with only one reference gene.

An example of the increased cell-to-cell variation in transcript levels in old mouse heart is shown in Fig. 2a for β -actin (*Actb*), β -2 microglobulin (*B2m*), lipoprotein lipase (*Lpl*) and cardiac α -actin (*Actc1*), normalized over glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*). In contrast, no significant cell-to-cell variation was observed when quantifying the three mitochondrial cytochrome *c* oxidase genes—*COX1*, *COX2* (also known as *mt-Co2*) and *COX3* (also known as *mt-Co3*)—with either of these as the reference gene

(Fig. 2b; Table 1). This is not unexpected, as these genes are transcribed as one long RNA molecule with many thousands of mitochondria present in each cardiomyocyte. When normalized against a nuclear gene, such as *Gapdh*, significant cell-to-cell variation in the relative quantities of each mitochondrial gene was always found (data not shown). These results underscore the stochasticity residing in nuclear genes and independently confirm the unbiased nature and reproducibility of the global mRNA amplification method.

We consider it unlikely that the observed cell-to-cell variation in relative gene expression and its increase with age are caused by the isolation procedure, because a doubling of collagenase concentration or incubation time resulted in the same range of cell-to-cell variation and a similar increase with age (data not shown). Nevertheless, we tested for a possible systematic effect on gene expression in the different cells by constructing heat plots for each group of cells—that is, from the three young and three old animals. The results show no specific correlation of the relative gene expression levels for any of the cells, allowing us to conclude that the observed variation is random and free of systematic bias (Fig. 2c).

Noise is inherent in the basic process of transcription, especially for genes expressed at low levels^{9,10}. We considered the possibility that alterations in the somatic genome—for example, as induced by reactive oxygen species (a likely causal factor in ageing)—could contribute to increased stochasticity of gene expression. To confirm that genotoxic treatment does lead to increased stochasticity of gene expression in cells, we treated cultured mouse embryonic fibroblasts (MEFs) at early passage with 0.1 mM hydrogen peroxide (H_2O_2), a known generator of oxidative damage. At this dose the survival rate is approximately 80%. At different time points after treatment, individual cells were obtained and global mRNA amplified, similar to the cardiomyocytes from young and old heart. For these cells, we assessed the mRNA levels of three housekeeping genes—that is, *Actb*, *B2m* and *Tuba6* with *Gapdh* as the normalizer. As compared with single-cell equivalents, untreated cells showed a trend ($P = 0.0847$) towards cell-to-cell variation in gene expression, similar to the situation in cardiomyocytes from young animals, albeit with less magnitude (Fig. 3a). At 48 h after the treatment with H_2O_2 , cell-to-cell variation in gene expression greatly increased. This is shown in Fig. 3a for *Actb* as an example, but a similar increase was observed for *B2m* and *Tuba6* (data not shown).

To investigate when this H_2O_2 -induced transcriptional noise occurred and how long it would last, we studied cell-to-cell variation in the expression of these three housekeeping genes at 6 h, 48 h and 9 days after treatment in several independent experiments. The results of one such experiment is shown in Fig. 3b. At 6 h after treatment, cell-to-cell variation in the relative expression level of none of these

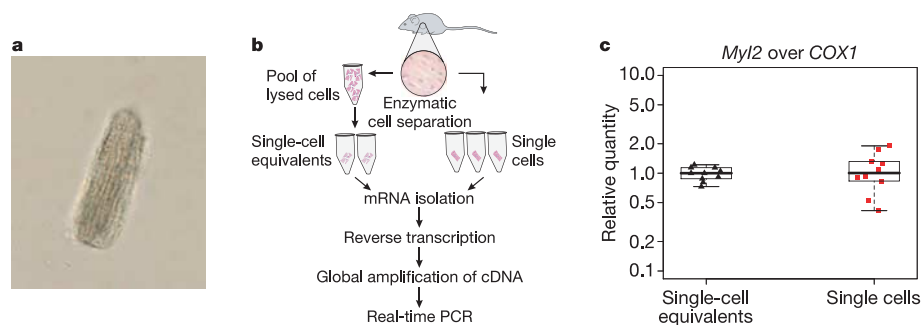


Figure 1 | Reproducibility of unbiased, global mRNA amplification of mouse cardiomyocytes. **a**, Typical morphology of cardiomyocytes after enzymatic dissociation. Original magnification, $\times 400$. **b**, Schematic representation of the strategy to validate cell-to-cell variation in gene expression as compared with the experimental variation in repeated global mRNA amplification. The single-cell equivalents were made from a pool of lysed cardiomyocytes. Fifteen single cells and ten single-cell equivalents were separately subjected

to mRNA isolation, reverse transcription and global cDNA amplification. **c**, Relative expression of the *Myl2* gene (over *COX1*). Significantly increased variability was observed among the single cells as compared with the single-cell equivalents ($P = 0.0026$). Boxes in the box plots indicate the interquartile range (IQR) with the median; the whiskers indicate $1.5\times$ the IQR.

genes was increased. In fact, untreated cells at this early stage sometimes showed a somewhat higher variation, for which we have no immediate explanation. This result strongly suggests that the increased transcriptional noise induced by H_2O_2 is not due to direct chemical damage of DNA, proteins or lipids—the effects of which are likely to appear almost immediately.

At 48 h after treatment, the increased stochasticity observed in the earlier experiment was reproduced. As documented by numerous

investigators, H_2O_2 induces replicative senescence in primary mouse or human fibroblasts¹¹. At 48 h after treatment, the cell populations started to show an increased number of senescent cells, characterized by their flattened and enlarged morphology (data not shown). During this period, when the cells were maintained in the same culture plates, cell death was not apparent, and at 9 days after treatment almost all cells were senescent. At that time, the increased cell-to-cell variation in gene expression among the H_2O_2 -treated populations was still highly significant (Fig. 3b). In Fig. 3c, we plotted the coefficients of variation at the three time points as averages of 2–4 independent experiments for the three genes together. The kinetics of the process indicates an increase at 48 h, after which time the transcriptional noise seemed to remain stable. On the basis of these results, we conclude that H_2O_2 -induced transcriptional noise is a relatively late event, which is virtually absent during the first hours after treatment. However, at least in a population of low cell-turn-over, the induced transcriptional noise is persistent, although a modest decline at 9 days cannot be ruled out.

As the H_2O_2 -induced transcriptional noise is unlikely to be caused by an early event, such as DNA damage or damage to proteins or

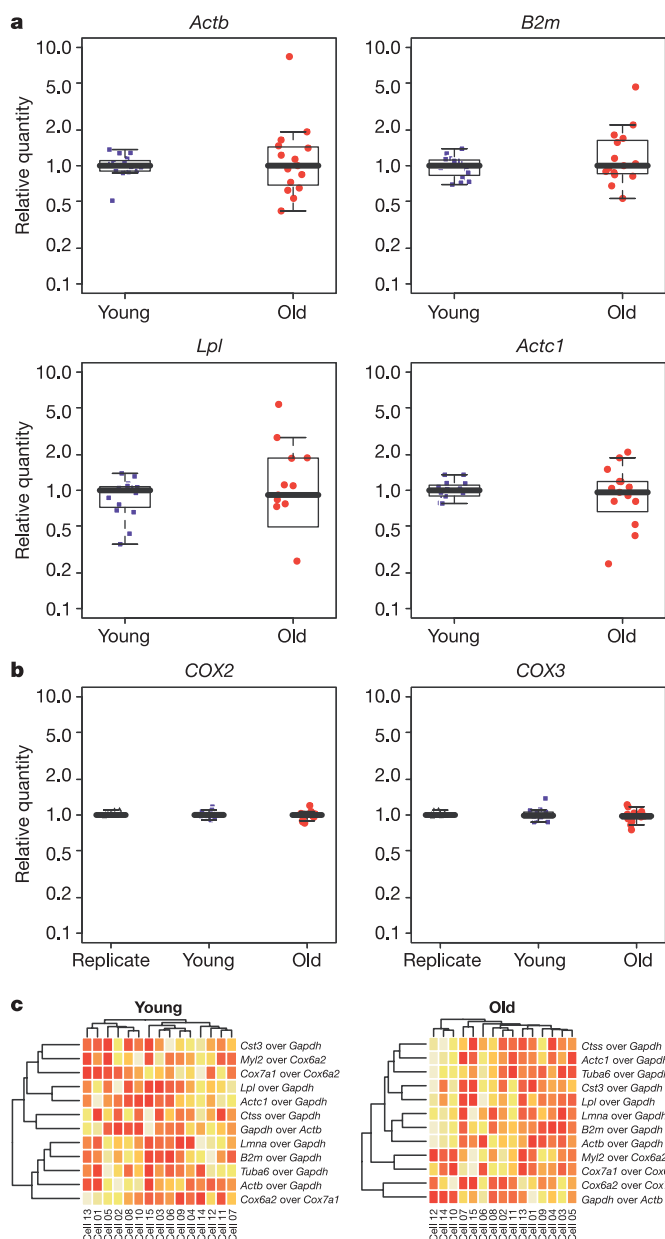


Figure 2 | Increased cell-to-cell variation in gene expression among cardiomyocytes from the heart of old as compared with young mice. **a**, Examples of four genes showing statistically significant cell-to-cell variability in expression (*Actb*, *B2m*, *Lpl* and *Actc1*; normalized over *Gapdh*). *P*-values can be found in Table 1. Box plots are as in Fig. 1. **b**, Expression of mitochondrial cytochrome *c* oxidase genes (normalized over *COX1*) in the same cells showed no statistically significant increase in stochastic variation. Ten replicates from the same template were analysed along with the young and old cells to determine the real-time PCR error. Box plots are as in Fig. 1. **c**, Representative heat maps of one young and one old animal revealed no significant pattern of change in $\Delta\Delta C_T$ (ranked for each gene) among the cells (red = low $\Delta\Delta C_T$ levels; white = high $\Delta\Delta C_T$ levels), indicating that the observed cell-to-cell variation is random.

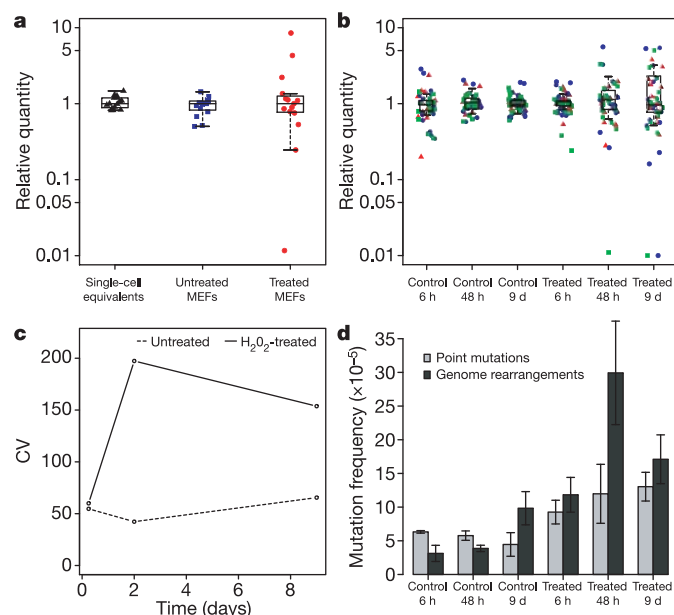


Figure 3 | H_2O_2 treatment of MEFs increases cell-to-cell variation in gene expression in parallel to the induction of genome rearrangements. **a**, Cell-to-cell variation in relative expression of *Actb* (normalized over *Gapdh*) among MEF single-cell equivalents, untreated MEFs, and MEFs at 48 h after treatment with 0.1 mM H_2O_2 . The variation among the treated cells is significantly greater than among the untreated cells ($P < 0.0001$). Box plots are as in Fig. 1. **b**, Cell-to-cell variation in relative expression of *Actb* (red), *B2m* (blue) and *Tuba6* (green) (normalized over *Gapdh*) at 6 h, 48 h and 9 days after treatment with H_2O_2 . Both at 48 h and at 9 days, cell-to-cell variation in relative expression of the three genes was significantly greater than in the controls ($P < 0.0001$ in both cases). At 6 h, variation among the untreated, control cells was somewhat higher than among the treated cells ($P = 0.0382$), but only in this experiment. Box plots are as in Fig. 1. **c**, Coefficient of variation (CV) calculated for all relative quantities of *Actb*, *B2m* and *Tuba6* at each time point and treatment condition. The 6 h, 48 h and 9 day time points are based on 2, 4 and 3 independent experiments, respectively. Each experiment consists of 11–15 single-cell determinations for each gene under each treatment condition. **d**, *lacZ* mutation frequencies at 6 h, 48 h and 9 days after H_2O_2 treatment as compared with control populations. Mutation frequencies are averages from three independent determinations from each of two parallel experiments. The subdivision into point mutations (grey bars) and genome rearrangements (black bars) was made on the basis of restriction enzyme analysis of 48 mutants taken from each experiment. Error bars indicate s.d.

lipids, we analysed mutations at a *lacZ* reporter gene as a relatively late marker for genomic instability. The *lacZ* reporter genes were present in the C57BL/6 mouse line from which the cells had been derived¹². The results of this analysis indicate that although some mutations were already present at 6 h after treatment, their frequency increased substantially at 48 h and declined somewhat at 9 days. Most of these mutations were large genome rearrangements, as indicated by restriction analysis of the mutant *lacZ* plasmids recovered from the cells (Fig. 3d). Hence, the kinetics of H₂O₂-induced mutations and transcriptional noise are quite similar, though not exactly the same.

Although it is tempting to suggest that the accumulation of genome rearrangements, which we previously demonstrated to accumulate with age in the mouse heart^{3,4}, is a cause of increased stochasticity of gene expression in ageing—for example, through gene dose and/or position effects—random changes in DNA methylation or chromatin remodelling could contribute as well. However, other cellular changes unrelated to genome alterations cannot be excluded at this stage. Stochastic changes in gene expression have been considered as a possible cause of ageing since Orgel proposed his ‘error catastrophe’ theory¹³, and the role of chance events in ageing is now increasingly appreciated^{14–16}. Our present results do not allow any further conclusions with respect to increased transcriptional noise as a mechanism of ageing. However, stochasticity is ubiquitous in biological systems and increased noise in gene expression has been implicated in reduced organismal fitness¹⁷. It is conceivable that a variety of persistent forms of damage to biological macromolecules could initiate a gradual increase in transcriptional noise introducing phenotypic variation among cells. Such variation could generally become detrimental to normal cell functioning, which would explain many aetiological characteristics of the ageing process, most notably the highly variable, progressive decline in organ function.

METHODS

See Supplementary Methods for detailed experimental procedures.

Animals and tissue collection. Young (6 months) and old (27 months) BALB/c mice were obtained from the National Institute on Aging and used after two weeks. Animals were euthanized by CO₂ and the hearts perfused with PBS and excised. The ventricular parts were subsequently dissected and either flash frozen and stored at -80°C or immediately used for single-cell isolation.

Isolation of cardiomyocytes. The ventricular tissue was minced and incubated for 30 min at room temperature ($\sim 20^{\circ}\text{C}$) under gentle shaking in collagenase in Krebs–Henseleit buffer. Single cells were collected under an inverted microscope by hand-held capillaries, deposited in PCR tubes and immediately frozen on dry ice and stored at -80°C until needed. The viability of the cardiomyocytes was determined by Trypan-blue dye exclusion and found to be over 90% for both young and old mice.

Isolation of mouse embryonic fibroblasts and treatment with H₂O₂. Mouse embryonic fibroblasts (MEFs) were isolated from embryonic day 13.5 embryos of C57BL/6 mice harbouring multiple copies of chromosomally integrated plasmids containing the *lacZ* mutational reporter gene, as described¹⁸. For each experiment, a total of 1 million cells were seeded in a 10-cm culture dish and treated with 0.1 mM H₂O₂ (Sigma) in medium without serum for 2 h at 37°C .

Global mRNA amplification. mRNA isolation and global amplification of first-strand cDNA was carried out according to a modified version of the procedure described in ref. 7. For each young versus old animal pair, cells were processed simultaneously using the same reagents.

Real-time PCR. Gene expression levels were quantified by real-time PCR, using the ABI Prism 7900HT system (Applied Biosystems). A list of the panel of genes tested along with the primer sequences used is provided in Supplementary Table 1.

Detection and analysis of genomic mutations. DNA was extracted from treated and control cells, and mutant frequencies and spectra were determined as described¹⁹. For each sample, at least 48 mutants were characterized.

Statistical analysis. For each target gene studied, the expression level relative to that of the reference gene and calibrator ($\Delta\Delta C_{\text{T}}$; see Supplementary Methods for further definition) was modelled as a linear function of mouse age. Bartlett’s tests were used to determine if the variance of young and old mice differed significantly.

Received 21 December 2005; accepted 28 April 2006.

- Vijg, J. & Dollé, M. E. Large genome rearrangements as a primary cause of aging. *Mech. Ageing Dev.* **123**, 907–915 (2002).
- Kirkwood, T. B. Understanding the odd science of aging. *Cell* **120**, 437–447 (2005).
- Dollé, M. E., Snyder, W. K., Gossen, J. A., Lohman, P. H. & Vijg, J. Distinct spectra of somatic mutations accumulated with age in mouse heart and small intestine. *Proc. Natl Acad. Sci. USA* **97**, 8403–8408 (2000).
- Dollé, M. E. & Vijg, J. Genome dynamics in aging mice. *Genome Res.* **12**, 1732–1738 (2002).
- Levsky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. Single-cell gene expression profiling. *Science* **297**, 836–840 (2002).
- Levsky, J. M. & Singer, R. H. Gene expression and the myth of the average cell. *Trends Cell Biol.* **13**, 4–6 (2003).
- Klein, C. A. *et al.* Combined transcriptome and genome analysis of single micrometastatic cells. *Nature Biotechnol.* **20**, 387–392 (2002).
- Kang, P. M., Haunzetter, A., Aoki, H., Ushuva, A. & Izumo, S. Morphological and molecular characterization of adult cardiomyocyte apoptosis during hypoxia and reoxygenation. *Circ. Res.* **87**, 118–125 (2000).
- Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6**, 451–464 (2005).
- Becskei, A., Kaufmann, B. B. & van Oudenaarden, A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature Genet.* **37**, 937–944 (2005).
- Parrinello, S. *et al.* Oxygen sensitivity severely limits the replicative lifespan of murine fibroblasts. *Nature Cell Biol.* **5**, 741–747 (2003).
- Boerigter, M. E., Dollé, M. E., Martus, H. J., Gossen, J. A. & Vijg, J. Plasmid-based transgenic mouse model for studying *in vivo* mutations. *Nature* **377**, 657–659 (1995).
- Orgel, L. E. The maintenance of the accuracy of protein synthesis and its relevance to ageing. *Proc. Natl Acad. Sci. USA* **49**, 517–521 (1963).
- Finch, C. E. & Kirkwood, T. B. *Chance, Development, and Aging* (Oxford Univ. Press, New York, 2000).
- Rea, S. L., Wu, D., Cypser, J. R., Vaupel, J. W. & Johnson, T. E. A stress-sensitive reporter predicts longevity in isogenic populations of *Caenorhabditis elegans*. *Nature Genet.* **37**, 894–898 (2005).
- Golden, T. R. & Melov, S. Microarray analysis of gene expression with age in individual nematodes. *Ageing Cell* **3**, 111–124 (2004).
- Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J. & Eisen, M. B. Noise minimization in eukaryotic gene expression. *PLoS Biol.* **2**, e137 (2004).
- Busuttill, R. A., Rubio, M., Dollé, M. E., Campisi, J. & Vijg, J. Oxygen accelerates the accumulation of mutations during the senescence and immortalization of murine cells in culture. *Ageing Cell* **2**, 287–294 (2003).
- Dollé, M. E., Martus, H. J., Gossen, J. A., Boerigter, M. E. & Vijg, J. Evaluation of a plasmid-based transgenic mouse model for detecting *in vivo* mutations. *Mutagenesis* **11**, 111–118 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by an NIH grant to J.V. and a BioFuture grant from the German Federal Ministry for Education and Science to C.A.K. We thank K. Khrapko and R. Beems for advice and useful suggestions.

Author Contributions J.V. conceived, designed and supervised the study (initially with M.E.T.D.), obtained the funding and took the primary role in writing the paper. R.B. designed and performed the experiments, assisted by K.A.R., A.D.D. and R.A.B. R.B.C., G.B.C. and B.H.P. performed statistical analyses and made the figures. C.H.H. and C.A.K. provided the modified protocol for global mRNA amplification.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.V. (jjvig@buckinstitute.org).

Smad4 signalling in T cells is required for suppression of gastrointestinal cancer

Byung-Gyu Kim¹, Cuiling Li², Wenhui Qiao², Mizuko Mamura¹, Barbara Kasperczak³, Miriam Anver³, Lawrence Wolfrum¹, Suntaek Hong¹, Elizabeth Mushinski⁴, Michael Potter⁴, Seong-Jin Kim¹, Xin-Yuan Fu⁵, Chuxia Deng² & John J. Letterio¹

SMAD4 (MAD homologue 4 (*Drosophila*)), also known as **DPC4** (deleted in pancreatic cancer), is a tumour suppressor gene that encodes a central mediator of transforming growth factor- β signalling^{1–4}. Germline mutations in **SMAD4** are found in over 50% of patients with familial juvenile polyposis, an autosomal dominant disorder characterized by predisposition to hamartomatous polyps and gastrointestinal cancer^{5,6}. Dense inflammatory cell infiltrates underlay grossly normal appearing, non-polypoid colonic and gastric mucosa of patients with familial juvenile polyposis⁷. This prominent stromal component suggests that loss of **SMAD4**-dependent signalling in cells within the epithelial microenvironment has an important role in the evolution of intestinal tumorigenesis in this syndrome. Here we show that selective loss of **Smad4**-dependent signalling in T cells leads to spontaneous epithelial cancers throughout the gastrointestinal tract in mice, whereas epithelial-specific deletion of the *Smad4* gene does not. Tumours arising within the colon, rectum, duodenum, stomach and oral cavity are stroma-rich with dense plasma cell infiltrates. *Smad4*^{−/−} T cells produce abundant T_H2-type cytokines including interleukin (IL)-5, IL-6 and IL-13, known mediators of plasma cell and stromal expansion. The results support the concept that cancer, as an outcome, reflects the loss of the normal communication between the cellular constituents of a given organ⁸, and indicate that *Smad4*-deficient T cells ultimately send the wrong message to their stromal and epithelial neighbours.

We proposed that *Smad4*-dependent signalling in T lymphocytes is required to maintain immune homeostasis within the intestinal environment, and that specific disruption of *Smad4* within the T lineage in mice would be sufficient to induce the formation of hamartomatous lesions and cancer within the gastrointestinal mucosa. To test this hypothesis, we used conditional gene targeting to generate mouse models in which the *Smad4* gene is either specifically deleted in T cells or broadly within epithelia, including the intestinal mucosa. This strategy provides the advantage that one can clearly assess the contribution of a disruption in *Smad4* gene expression within one lineage, while *Smad4* expression remains intact in the other. To inactivate specifically *Smad4* in T cells, we crossed mice homozygous for a *Smad4* conditional allele (*Smad4*^{col/co})⁹ with mice expressing a transgene encoding a Cre recombinase driven either by the Lck promoter (*Smad4*^{col/co;Lck-cre})¹⁰ or the CD4 promoter (*Smad4*^{col/co;CD4-cre})¹¹ (Fig. 1a–d). Cre-mediated excision of exon 8 of the *Smad4* gene was detected by polymerase chain reaction (PCR) in T cells of either *Smad4*^{col/co;Lck-cre} or *Smad4*^{col/co;CD4-cre}

mice, but not in T cells of *Smad4*^{+/+} mice (Fig. 1c, top panel). Analysis by western blot shows the absence of *Smad4* protein in T cells purified from either *Smad4*^{col/co;Lck-cre} or *Smad4*^{col/co;CD4-cre} mice (Fig. 1c, bottom panel). Immunohistochemical staining through sections of intestine of either *Smad4*^{col/co;Lck-cre} or *Smad4*^{col/co;CD4-cre} mice clearly demonstrates normal expression of *Smad4* in epithelial cells, and in plasma cells within the intimal layer (Fig. 1d). Two additional models were generated in which epithelial-specific promoters (either mouse mammary tumour virus (MMTV)-Cre or transthyretin-Cre) were used to inactivate directly *Smad4* expression in intestinal epithelia (Fig. 1e–j)^{12,13}, leaving lymphocyte expression of *Smad4* intact.

All *Smad4*^{col/co;Lck-cre} and *Smad4*^{col/co;CD4-cre} mice develop normally, but begin to experience weight loss at 3 months of age (Fig. 2a, b), and eventually display clinical features of systemic illness (lethargy, hunched posture, dishevelled fur) (Fig. 2a). Survival of *Smad4*^{col/co;Lck-cre} and *Smad4*^{col/co;CD4-cre} mice is significantly shortened when compared to wild-type or heterozygous littermates (Fig. 2c, d). Necropsy studies revealed significant gastrointestinal pathology as a cause of clinical symptoms (Fig. 2e–i). Gross enlargement of both the small and large intestines is present in all *Smad4*^{col/co;Lck-cre} and *Smad4*^{col/co;CD4-cre} mice over 9 months of age at necropsy, and rectal prolapse was observed in 25% (4 out of 16) of mice deficient for T-cell expression of *Smad4* (Fig. 2e). The intestinal mucosal surface was grossly thickened, with both linear sessile and peduncular polyps visible within the lumen of the intestine (Fig. 2f, g). Histological analysis through cross-sections of the intestine showed effacement of the normal villus architecture by a stroma-rich mixed mononuclear cell infiltrate (Fig. 2h, i).

The histopathology of the intestines of *Smad4*^{col/co;Lck-cre} and *Smad4*^{col/co;CD4-cre} mice is, in many respects, a phenocopy of the pathology found in familial juvenile polyposis (FJP). The large and small intestines are diffusely involved, containing cystic spaces lined by columnar epithelial cells, surrounded by abundant stroma (Fig. 3a). This stroma-rich intimal layer contains a dense plasma cell infiltrate (Fig. 3b), clearly demonstrated by immunostaining for κ light chain (Supplementary Fig. 2a, b). Plasma cell infiltrates are IgA-producing, with abnormal accumulation of IgA⁺ plasma cells in places such as the base of the glandular stomach (Fig. 3c), and a massive increase in the levels of IgA in the serum of *Smad4*^{col/co;Lck-cre} and *Smad4*^{col/co;CD4-cre} mice. Circulating IgA levels were greater than 2,500 mg dl^{−1} (normal levels are <220 mg dl^{−1}) (Fig. 3d).

Plasma cell differentiation and expansion within the intestinal environment is regulated by an array of T_H2 cytokines derived from

¹Laboratory of Cell Regulation and Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

²Genetics of Disease and Development Branch, National Institutes of Diabetes & Digestive & Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

³SAIC, National Cancer Institute-Frederick, PO Box B, Frederick, Maryland 21702, USA. ⁴Laboratory of Genetics, Center for Cancer Research, The National Institutes of Health, Bethesda, Maryland 20892, USA. ⁵Department of Microbiology and Immunology, Indiana School of Medicine, Indianapolis, Indiana 46202, USA.

CD4⁺ T-helper cells. Transforming growth factor- β (TGF- β) signalling has a major role in regulating T-cell activation, differentiation and cytokine production^{14,15}. We proposed that loss of Smad4-dependent signalling in T cells would potentiate their activation and lead to increased production of cytokines, including the T_H2 cytokines IL-5 and IL-6. To test this hypothesis, naive T cells were activated *in vitro* under neutral priming conditions and supernatants were analysed for cytokine production by a cytokine protein array. Smad4-deficient T cells were skewed towards a T_H2 phenotype, with high levels of IL-4, IL-5, IL-6 and IL-13 in supernatants of T cells isolated from *Smad4*^{col/co;Lck-cre} and *Smad4*^{col/co;CD4-cre} mice, when compared to production by T cells of *Smad4*^{+/+} mice (Fig. 3e).

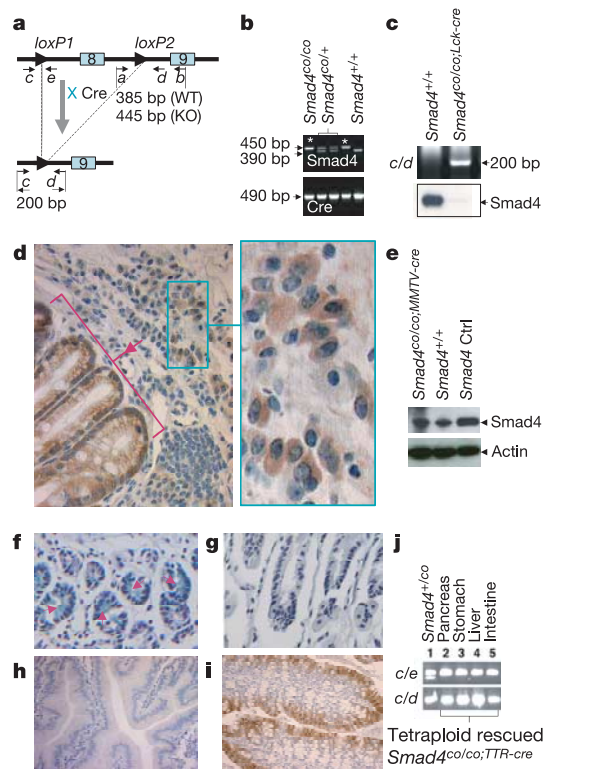


Figure 1 | Conditional deletion of exon 8 of the *Smad4* gene in either epithelia or T lymphocytes. **a**, Cre-recombinase-mediated excision of *Smad4* conditional allele with loxP sites (*Smad4*^{col/co}). **b**, Genotyping shows wild-type and conditional alleles in lines expressing the Lck-Cre recombinase transgene (asterisk indicates homozygote for conditional allele). **c**, Detection of recombined allele in thymocytes of *Smad4*^{col/co;Lck-cre} mice using primers *c/d* (top panel); loss of Smad4 protein by western blot (bottom panel). **d**, Histochemical analysis of *Smad4*^{col/co;Lck-cre} mice shows expression of Smad4 in intestinal epithelia and plasma cells within the intimal layer (original magnification $\times 100$; plasma cells in panel outlined in blue, $\times 400$). **e**, Western blot analysis shows equivalent Smad4 expression in purified T cells of wild-type and *Smad4*^{col/co;MMTV-cre} mice. **f, g**, X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside) expression in mice generated by crossing MMTV-Cre transgenic mice with a lacZ conditional allele. Blue staining (**f**, red arrows) indicates Cre-recombinase expression in intestinal epithelial cells; intestine in the control lacking the reporter allele (**g**) is negative. **h, i**, Immunostaining shows the absence of Smad4 expression in intestinal sections of *Smad4*^{col/co;MMTV-cre} mice (**h**) and its presence in control mice (**i**) ($\times 100$ original magnification for **f–i**). **j**, Cre recombinase expressed under the control of a murine TTR promoter induces *Smad4* deletion in multiple tissues (PCR analysis). The top panel shows both wild type (+) and conditional (co) *Smad4* alleles in a heterozygote (lane 1), whereas the single band generated with primers *c/e* (lanes 2–5) indicates that mice derived by the technique of tetraploid rescue are homozygous for the conditional allele (that is, the wild-type configuration is not found when using identical primers). The recombined allele is detected in all tissues where the conditional allele is expressed (bottom panel, primer pair *c/d*).

IL-5 and IL-6 have important roles in inducing proliferation of pre-activated B cells and plasma cell differentiation, with IL-5 as the most potent cytokine for stimulating production and secretion of IgA. IL-13 has been implicated in mediating the induction of matrix proteins and fibrosis in allergic disorders dominated by T_H2 cytokine production¹⁶. These cytokines were also significantly elevated in the sera of *Smad4*^{col/co;Lck-cre} and *Smad4*^{col/co;CD4-cre} mice, demonstrating that this production of T_H2 cytokines occurs *in vivo* (shown for IL-6 in Fig. 3f).

These data implicate defects in SMAD4-mediated signalling in T cells in the pathogenesis of FJP. This association is strengthened by the spontaneous progression to epithelial cancer in the gastrointestinal tract of both *Smad4*^{col/co;Lck-cre} and *Smad4*^{col/co;CD4-cre} mice. Epithelial cancers were found in 94% (16 out of 17) of *Smad4*^{col/co;Lck-cre} mice and 100% (11 out of 11) of *Smad4*^{col/co;CD4-cre} mice necropsied between 9 and 16 months of age (Fig. 4). The histopathology included epithelial lesions with atypia, adenomas (Fig. 4a) and invasive carcinomas with metastases (Fig. 4c–f). Importantly, an evaluation of the expression of Smad4 (Fig. 4g, h) and of phosphorylated Smad2 (Fig. 4i, j) shows that Smad-dependent

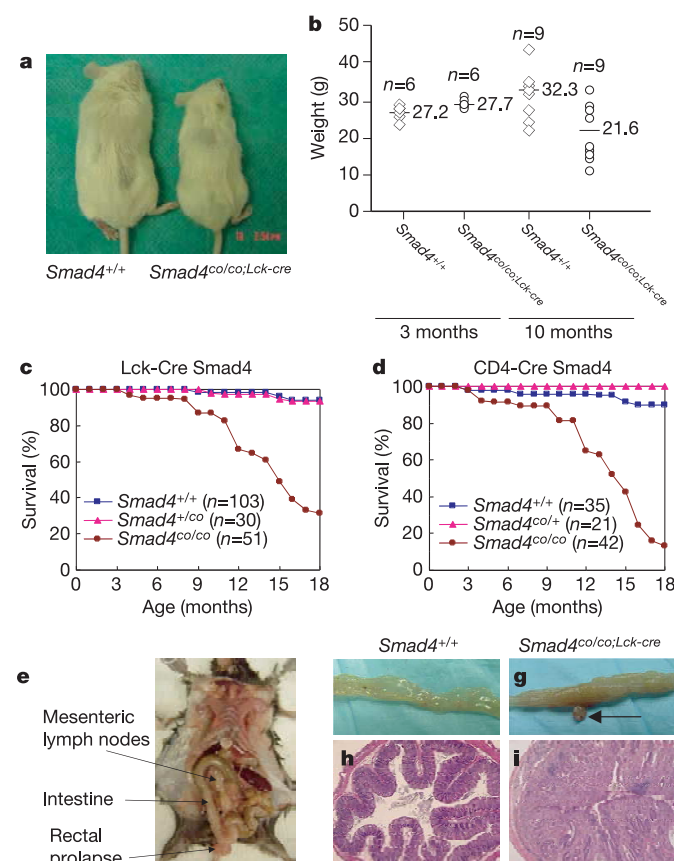


Figure 2 | T-cell-restricted deletion of the *Smad4* gene in mice leads to gastrointestinal pathology and decreased survival. **a**, Representative photograph of *Smad4*^{+/+} and *Smad4*^{col/co;Lck-cre} mice at 10 months of age shows decreased body mass and dishevelled appearance of *Smad4*^{col/co;Lck-cre} mice. **b**, Both *Smad4*^{col/co;CD4-cre} mice (not shown) and *Smad4*^{col/co;Lck-cre} mice develop symptoms after 3 months. **c, d**, Survival curves of *Smad4*^{+/+} and *Smad4*^{col/co;CD4-cre} mice demonstrate lethality associated with T-cell-restricted deletion of *Smad4*. **e**, Thickened large intestine and rectal prolapse is common to both *Smad4*^{col/co;Lck-cre} and *Smad4*^{col/co;CD4-cre} lines. **f, g**, Colons of *Smad4*^{+/+} and *Smad4*^{col/co;Lck-cre} mice (opened for inspection of mucosal surfaces). In *SMAD4*^{col/co;Lck-cre} mice (**g**) there is thickening of the colon, with diffuse linear and pedunculated polyps (arrow). **h, i**, Cross-sections of intestine stained by haematoxylin and eosin reveal effacement of villus architecture and accumulation of stromal elements in *Smad4*^{col/co;Lck-cre} mice. Original magnification in **h** and **i**, $\times 50$.

TGF- β signalling remains intact in the epithelial compartment. The high frequency of squamous cell carcinoma in the oral cavity of both *Smad4^{col/co};Lck-cre* and *Smad4^{col/co};CD4-cre* mice is the single feature that seems to distinguish the mouse model from the syndrome of FJP in humans (Fig. 4k, l).

To support further the hypothesis that a T-cell defect in SMAD4-dependent signalling contributes to the pathogenesis of FJP in humans, none of the pathologies described above was observed in two mouse models in which the deletion of the *Smad4* gene is restricted to the epithelial lineage. In one model, Cre-mediated

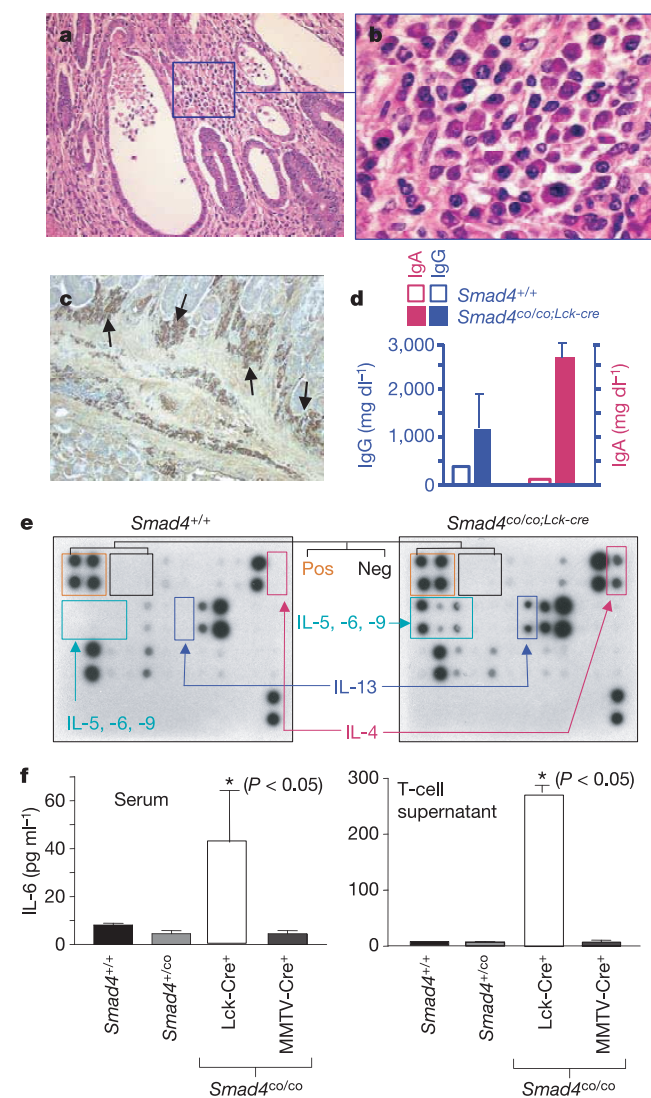


Figure 3 | T_H2 cytokines and plasma cells in mice with T-cell-restricted *Smad4* gene deletion. **a**, Haematoxylin and eosin-stained sections of a *Smad4^{col/co};Lck-cre* mouse intestine show accumulation of stroma surrounding cystic structures lined by columnar epithelia. Original magnification $\times 200$. **b**, Stroma are infiltrated by dense infiltrates of plasma cells. Original magnification $\times 400$. **c**, Plasma cells at the base of the glandular stomach are IgA-producing (immunohistochemical staining). Original magnification $\times 200$. **d**, Increased IgA levels (red) in sera of both *Smad4^{col/co};Lck-cre* (filled bars) and *Smad4^{col/co};CD4-cre* (not shown) mice. IgG (blue) is also increased relative to control *Smad4^{+/+}* mice (open bars). **e**, Cytokine protein array demonstrates T_H2 cytokine production by T cells lacking *Smad4* (supernatants collected 48 h after activation in culture). Both *Smad4^{col/co};Lck-cre* and *Smad4^{col/co};CD4-cre* mice show increased production of IL-4, IL-5, IL-6, IL-9 and IL-13. All results were confirmed by ELISA (shown for IL-6 in **f**; error bars indicate s.d. of the mean, $n = 8$ for all groups except for MMTV-Cre⁺, where $n = 4$).

deletion of exon 8 of the *Smad4* gene was driven by the MMTV promoter. *Smad4^{col/co};MMTV-cre* mice develop squamous cell carcinomas in both the skin¹⁷ and the mammary gland between 5 and 16 months of age¹⁸. MMTV-Cre does not induce recombination and *Smad4* deletion in T cells, as purified T lymphocytes from spleen and lymph nodes of *SMAD4^{col/co};MMTV-cre* mice express *Smad4* protein at levels equivalent to those found in T cells of *Smad4^{+/+}* mice (Fig. 1e). Although MMTV-driven expression of Cre recombinase in the intestine effectively disrupts epithelial expression of *Smad4* (Fig. 1f, g), it does not lead to epithelial carcinoma anywhere in the gastrointestinal tract. In a second model, a murine transthyretin (TTR) promoter that is also expressed in the gastrointestinal tract successfully promotes *Smad4* gene deletion in several tissues¹³, including the stomach and intestine (Fig. 1j). Once again, disruption of *Smad4* expression within the epithelial compartment alone was not sufficient to produce spontaneous epithelial cancers in *Smad4^{col/co};TTR-cre* mice. We have also evaluated the effect of epithelial-specific expression of a dominant-negative *Smad4* transgene expressed under the control of the mouse intestinal trefoil

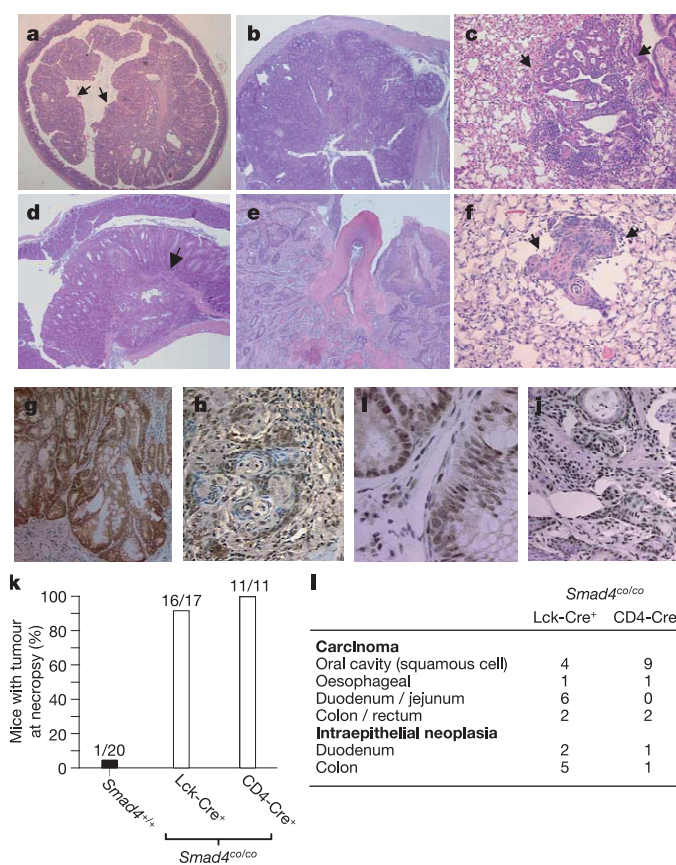


Figure 4 | Spontaneous gastrointestinal tumours induced by T-cell-restricted deletion of *Smad4*. **a–e**, Haematoxylin and eosin-stained tumours of *Smad4^{col/co};Lck-cre* and *Smad4^{col/co};CD4-cre* mice. **a**, Large intraluminal duodenal adenoma. **b**, Invasive colon carcinoma. **c**, Metastatic adenocarcinoma in lung (between arrows). **d**, Invasive rectal carcinoma (arrow indicates invasion below muscularis mucosa). **e**, Mandibular infiltration of a squamous cell carcinoma of the oral cavity (shown surrounding a tooth). **f**, Metastatic squamous cell carcinoma in lung (between arrows). **g**, **h**, Immunostaining of adenocarcinoma (**g**) and squamous cell carcinoma (**h**) indicates that *Smad4* expression is retained in tumour epithelia. **i**, **j**, Immunostaining for phosphorylated *Smad2* in colon (**i**) and squamous cell carcinoma (**j**) of *Smad4^{col/co};CD4-cre* mice shows nuclear accumulation of phosphorylated *Smad2* in epithelia, indicating that proximal TGF- β signalling remains intact in the epithelial compartment. **k**, **l**, Incidence and spectrum of gastrointestinal tumours. Original magnification: **a**, **b**, **d**, **e**, $\times 50$; **c**, **f**, **g–i**, $\times 200$.

factor (designated ITF-dnS4). This transgene encodes a radically truncated Smad4 (lacking the DNA-binding MH1 domain), which is strongly expressed in epithelia of the mouse stomach and intestine (Supplementary Fig. 3). Although this dominant-negative version of Smad4 effectively abrogates TGF- β -induced signalling via the Smad pathway¹⁹, *in vivo* expression of the dnS4 transgene in the intestinal epithelia neither adversely affects intestinal function nor leads to spontaneous tumorigenesis.

It is widely accepted that the development of carcinoma, the most common form of human cancer, is due to the accumulation of somatic mutations in epithelial cells. The TGF- β signalling pathway has an important role in this process, with mutations in the TGF- β type II receptor frequently found in intestinal epithelial cells of patients with colon cancer²⁰. However, the initiation and progression of carcinomas is also influenced by the tumour microenvironment, which includes extracellular matrix, blood vasculature, inflammatory cells and fibroblasts^{8,21}. The role of TGF- β in maintaining this epithelial 'landscape' is supported by the observation that disruption of TGF- β signalling in fibroblasts modulates the oncogenic potential of adjacent epithelia via a mechanism that involves the production of soluble factors that provide mitogenic signals in an abnormal paracrine signalling loop²².

The concept that SMAD4 might mediate tumour suppression by regulating cells within the tumour environment²³ has been questioned, in part, because of the finding of SMAD4 loss of heterozygosity (LOH) within the epithelial tumours of FJP patients²⁴. The data presented here suggest a need to reconcile these two models. The fact that SMAD4 LOH is observed in the tumour epithelia does not omit the possibility that alterations in Smad4-dependent signalling in other lineages might have a role, and possibly even select for LOH within the epithelial compartment. Although SMAD4 LOH has not been observed in the lymphoid lineage in patients with FJP, it is indeed possible that haploinsufficiency might exist for SMAD4 in the T lineage, and that this could work together with the complete loss of SMAD4 expression in the overlying epithelia. To address this hypothesis in our model, we have also examined mice in which only a single allele of *Smad4* is deleted exclusively in T cells (*Smad4*^{+/-co;Lck-cre} mice). All *Smad4*^{+/-co;Lck-cre} mice show a significant degree of plasma cell hyperplasia, particularly within the duodenum, which in some areas associates with the formation of polyps (Supplementary Fig. 4a–d) and with epithelial hyperplasia in the stomach. Furthermore, we have also analysed the expression of IL6 α (a component of the IL-6 receptor), and find IL6 α protein in the epithelial scrapings from intestines of mice with either one or both *Smad4* alleles deleted only in T cells (Supplementary Fig. 4e). IL6 α is not present in epithelial scrapings from the intestines of normal controls or of mice in which MMTV-Cre was used to mediate Smad4 excision. Ultimately, this question must be addressed by analysis of T-cell function in patients with FJP.

The role of the Smad family of signalling intermediates in the regulation of T-cell development and function has yet to be clearly defined. The data presented here indicate that Smad4-dependent signalling in T cells is essential for the maintenance of homeostasis within the gastrointestinal microenvironment, and that progression to epithelial cancer results directly from the restricted deletion of *Smad4* in the T-cell lineage. Although we have not demonstrated a requirement for IL-6 in the pathogenesis of tumours in this model, it is possible that the aberrant paracrine effects of IL-6 acting on resident B cells and epithelial cells underlie progression to cancer^{25,26,27}. In mice, expression of a constitutively active gp130 subunit of the IL-6 receptor promotes adenomatous gastric hyperplasia and abrogates the epithelial cell response to TGF- β by inducing the expression of Smad7, an inhibitor of TGF- β -induced Smad signalling²⁸. Disruption of TGF- β signalling in T cells (through transgenic expression of a dominant negative TGF- β type II receptor) has been shown to accelerate azoxymethane-induced colon carcinogenesis through a mechanism that links T-cell production of IL-6 to

hyper-activation of STAT3 in mucosal epithelia²⁹. Additional genetic studies in which we combine a Smad4 defect in T cells with a disruption of the IL-6 signalling system in epithelial cells may help to resolve this question.

The data presented here link the loss of Smad4-dependent signalling in T cells to an increased susceptibility to spontaneous gastrointestinal tumorigenesis, and have important implications for the management of patients with FJP. Therapeutic strategies that target the activation and function of T cells may represent an effective form of preventive therapy that will potentially delay or obviate the need for aggressive surgical intervention. Finally, whereas Smad4 can function as a classical tumour suppressor, directly suppressing transformation in a cell autonomous manner, it has a significant role in maintaining intestinal homeostasis, preventing cancer through control over the intestinal lymphocytes and their response to signals within the mucosal environment.

METHODS

Mice and genotyping. Mice (C57BL/6 $\times$ SvEv129 $\times$ FVB) homozygous for a conditional *Smad4* allele (*Smad4*^{co/co})⁹ were bred with the following transgenic mice: (1) Lck-Cre (Taconic) or CD4-Cre (Taconic) to generate *Smad4*^{co/co;Lck-cre} or *Smad4*^{co/co;CD4-cre} mice lacking Smad4 specifically in T cells; and (2) MMTV-Cre (refs 12, 18) or TTR-Cre (ref. 13 and X.-Y.F., unpublished data) to generate *Smad4*^{co/co;MMTV-cre} or *Smad4*^{co/co;TTR-cre} mice lacking Smad4 within intestinal epithelia. F₁ offspring heterozygous for the conditional allele (*Smad4*^{+/-co}) were inbred to generate homozygous *Smad4* conditional (*Smad4*^{co/co}) mice with Cre-recombinase transgenes and control mice (such as *Smad4*^{co/co}, *Smad4*^{+/-co}, *Smad4*^{+/-+}). Full details for genotyping of conditional and transgenic lines are available in Supplementary Methods. Protocols were approved by the Animal Care and Use Committee of the NCI. Animal procedures were conducted in compliance with National Institutes of Health Guidelines.

Cell purification and culture. Cell suspensions were prepared from spleens and lymph nodes as described in Supplementary Methods. Antibodies used in FACS analyses and purification of T cells were purchased from BD Pharmingen. For cytokine production measurements, 5×10^4 lymph node T cells were stimulated in 0.2 ml complete media per well in 96-well plates coated with $2 \mu\text{g ml}^{-1}$ anti-CD3 monoclonal antibody and $1 \mu\text{g ml}^{-1}$ anti-CD28 monoclonal antibody (Pharmingen). After 48 h, supernatants were assayed for cytokine levels by an enzyme-linked immunosorbent assay (R&D Systems) and a mouse cytokine antibody array (RayBiotech).

Histology and immunohistochemistry. Smad4 and phosphorylated Smad2 antibodies were purchased from Santa Cruz Biotechnology and Cell Signalling Technology. Antibodies to κ light chain and IgA were purchased from BD Transduction Laboratories. Details regarding tissue preparation and immunohistochemical procedures are available in Supplementary Methods.

Western blotting. Preparation of extracts and western blotting methods are available in Supplementary Information. Smad4 antibody (Santa Cruz) was visualized with HRP-conjugated goat anti-mouse IgG, followed by ECL detection (Pierce).

Received 9 February; accepted 28 April 2006.

1. Heldin, C. H., Miyazono, K. & ten Dijke, P. TGF- β signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**, 465–471 (1997).
2. Hahn, S. A. et al. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271**, 350–353 (1996).
3. Massague, J. TGF- β signal transduction. *Annu. Rev. Biochem.* **67**, 753–791 (1998).
4. Miyaki, M. & Kuroki, T. Role of Smad4 (DPC4) inactivation in human cancer. *Biochem. Biophys. Res. Commun.* **306**, 799–804 (2003).
5. Howe, J. R. et al. Common deletion of SMAD4 in juvenile polyposis is a mutational hotspot. *Am. J. Hum. Genet.* **70**, 1357–1362 (2002).
6. Howe, J. R. et al. Mutations in the SMAD4/DPC4 gene in juvenile polyposis. *Science* **280**, 1086–1088 (1998).
7. Subramony, C., Scott-Conner, C. E., Skelton, D. & Hall, T. J. Familial juvenile polyposis. Study of a kindred: evolution of polyps and relationship to gastrointestinal carcinoma. *Am. J. Clin. Pathol.* **102**, 91–97 (1994).
8. Bissell, M. J. & Labarge, M. A. Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment? *Cancer Cell* **7**, 17–23 (2005).
9. Yang, X., Li, C., Herrera, P. L. & Deng, C. X. Generation of Smad4/Dpc4 conditional knockout mice. *Genesis* **32**, 80–81 (2002).
10. Orban, P. C., Chui, D. & Marth, J. D. Tissue- and site-specific DNA recombination in transgenic mice. *Proc. Natl Acad. Sci. USA* **89**, 6861–6865 (1992).

11. Sawada, S., Scarborough, J. D., Killeen, N. & Littman, D. R. A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. *Cell* **77**, 917–929 (1994).
12. Wagner, K. U. *et al.* Cre-mediated gene deletion in the mammary gland. *Nucleic Acids Res.* **25**, 4323–4330 (1997).
13. Yan, C., Costa, R. H., Darnell, J. E. Jr, Chen, J. D. & Van Dyke, T. A. Distinct positive and negative elements control the limited hepatocyte and choroid plexus expression of transthyretin in transgenic mice. *EMBO J.* **9**, 869–878 (1990).
14. Letterio, J. J. & Roberts, A. B. Regulation of immune responses by TGF- β . *Annu. Rev. Immunol.* **16**, 137–161 (1998).
15. Gorelik, L. & Flavell, R. A. Transforming growth factor- β in T-cell biology. *Nature Rev. Immunol.* **2**, 46–53 (2002).
16. Wynn, T. A. IL-13 effector functions. *Annu. Rev. Immunol.* **21**, 425–456 (2003).
17. Yang, L. *et al.* Targeted disruption of Smad4 in mouse epidermis results in failure of hair follicle cycling and formation of skin tumors. *Cancer Res.* **65**, 8671–8678 (2005).
18. Li, W. *et al.* Squamous cell carcinoma and mammary abscess formation through squamous metaplasia in Smad4/Dpc4 conditional knockout mice. *Development* **130**, 6143–6153 (2003).
19. Tsuchida, K., Zhu, Y., Siva, S., Dunn, S. R. & Sharma, K. Role of Smad4 on TGF- β -induced extracellular matrix stimulation in mesangial cells. *Kidney Int.* **63**, 2000–2009 (2003).
20. Markowitz, S. *et al.* Inactivation of the type II TGF- β receptor in colon cancer cells with microsatellite instability. *Science* **268**, 1336–1338 (1995).
21. Bhowmick, N. A., Neilson, E. G. & Moses, H. L. Stromal fibroblasts in cancer initiation and progression. *Nature* **432**, 332–337 (2004).
22. Bhowmick, N. A. *et al.* TGF- β signalling in fibroblasts modulates the oncogenic potential of adjacent epithelia. *Science* **303**, 848–851 (2004).
23. Kinzler, K. W. & Vogelstein, B. Landscaping the cancer terrain. *Science* **280**, 1036–1037 (1998).
24. Woodford-Richens, K. *et al.* Allelic loss at SMAD4 in polyps from juvenile polyposis patients and use of fluorescence *in situ* hybridization to demonstrate clonal origin of the epithelium. *Cancer Res.* **60**, 2477–2482 (2000).
25. Becker, C. *et al.* IL-6 signalling promotes tumor growth in colorectal cancer. *Cell Cycle* **4**, 217–220 (2005).
26. Hodge, D. R., Hurt, E. M. & Farrar, W. L. The role of IL-6 and STAT3 in inflammation and cancer. *Eur. J. Cancer* **41**, 2502–2512 (2005).
27. Corvinus, F. M. *et al.* Persistent STAT3 activation in colon cancer is associated with enhanced cell proliferation and tumor growth. *Neoplasia* **7**, 545–555 (2005).
28. Jenkins, B. J. *et al.* Hyperactivation of Stat3 in gp130 mutant mice promotes gastric hyperproliferation and desensitizes TGF- β signaling. *Nature Med.* **11**, 845–852 (2005).
29. Becker, C. *et al.* TGF- β suppresses tumor progression in colon cancer by inhibition of IL-6 trans-signaling. *Immunity* **21**, 491–501 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank A. B. Roberts for her thoughtful reading of this manuscript and considerate support of this research.

Author Contributions B.-G.K. and J.J.L. are responsible for developing the hypothesis, planning and executing experiments, including analysis of models with T-cell-specific deletion of *Smad4*, and for preparation of the manuscript. C.L., W.Q., X.-Y.F. and C.D. are responsible for the development of models with epithelial-specific deletion of the *Smad4* gene. S.-J.K. and S.H. developed and analysed the ITF-dnS4 transgenic model. M.M. established mouse colonies with T-cell-restricted deletion of *Smad4*. E.M. and M.P. provided immunohistochemical analysis of plasma cell infiltrates. M.A. provided analysis of histopathology and B.K. developed immunohistochemical procedures for *Smad4*. L.W. contributed in the preparation of the manuscript and in the evaluation of the data.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.J.L. (john.letterio@cwru.edu).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Medical students rarely come into contact with the world of basic research. But they should, argues a recent report from the Association of American Medical Colleges (AAMC). It would help them to turn such studies into better patient care, the AAMC argues, and would help them in their professional life as doctors to play a more effective role in research. The report, *Promoting Translational and Clinical Science: The Critical Role of Medical Schools and Teaching Hospitals*, recommends ways in which medical schools could produce a greater number of translational and clinical physician-scientists.

The first step, says the AAMC, is for future doctors — both as students and residents — to learn about basic, clinical and translational research as part of their course. Second, training opportunities need to be adjusted to include advanced degree courses, allowing potential physician-scientists to undertake research and receive appropriate mentoring. And finally, the career system needs to be altered so that it encourages and rewards research. If young physician-scientists are unlikely to be recognized for their contributions to clinical research or published papers, they will have little incentive to embrace research — especially when many young doctors are under pressure to see more patients.

Although these recommendations are specific to physician-scientists, the problems likely to arise if they are implemented are relevant to the wider scientific community. The first step is possible as it simply means changes to curricula at individual institutions. The second is harder because it will need senior scientists from medicine and beyond to buy into the idea — it is difficult to mandate better mentoring. But the third point is daunting, because it requires changes that are greater than individual scientists or institutions can bring to bear.

In all cases, policy-makers need to ask themselves how committed they are to the changes. But young scientists-to-be can help: by gravitating towards the best programmes, and by pushing for reforms when training and rhetoric don't match.

Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe

Editor: Paul Smaglik

Assistant Editor: Gene Russo

European Head Office, London

The Macmillan Building,
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)
Natureevents: Matt Betteridge (4015)

UK/RoW/Ireland:

Nils Moeller (4953)

Scandinavia/Spain/Portugal:

Claudia Paulsen Young (4973)

France/Switzerland/Belgium/Italy:

Maria Ceraolo (4994)

Germany/Austria/The Netherlands:

Reya Silao (4970)

Business Development Manager:

Amelie Pequignot (4974)

Advertising Production Manager:

Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Catherine Alexander

European Satellite Office

Germany:

Patrick Phelan

Tel: +49 89 54 90 57 11

Fax: +49 89 54 90 57 20

e-mail: p.phelan@nature.com

US Head Office, New York

75 Varick Street,
9th Floor, New York,
NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:

Ayako Watanabe

e-mail: a.watanabe@natureasia.com

Rules rule

Regulatory affairs is a young profession that's already making its mark in the world of drug development, where one false move can bring years of research to an unwelcome end. If your skills include communication and leadership, it may be for you, says **Hannah Hoag**.

Last year, global spending on prescription drugs exceeded US\$600 billion. Topping the sales list were drugs to combat high cholesterol levels, heartburn, schizophrenia and asthma. But regardless of their sales strength, every drug had one thing in common: each had been scrutinized by regulatory-affairs professionals before reaching the pharmacy shelves. Employed by the drug companies and contract research organizations, these experts assemble huge dossiers of data and oversee the mountain of paperwork required by the regulatory agencies that oversee the industry.

"Our basic responsibility is to interpret the regulations and guidelines of the different regulatory agencies," says Ken White, vice-president of regulatory affairs at ICOS in Bothell, Washington.

Regulatory-affairs professionals are an essential part of the drug-development process, and jobs in the field typically crop up wherever there is healthy investment in R&D. These professionals have a vast scope. They prepare, submit and monitor submissions to regulatory agencies, and get involved in manufacturing, labelling, marketing and post-marketing activities. Increasingly, their involvement begins during the R&D phase, where they help smooth the transition from drug development to clinical trials.

Once a company identifies a promising compound, a regulatory-affairs team begins drawing up the plan that will guide the compound through development, clinical trials and onto the shelves. "We work with the scientific team during the preclinical phase to make sure that we



have the right data," says Lisa Skeens, vice-president of regulatory affairs at Baxter in Deerfield, Illinois. "We are the liaison between the regulatory authorities and the company, the experts in understanding the regulations and the guidelines. We know through experience what the global regulatory agencies expect."

Before starting clinical trials, which involve testing the compound in people, companies must submit an Investigational New Drug (IND) application in the United States, or in Europe apply for a Clinical Trial Authorization. These provide evidence that a compound is both safe and has some beneficial activity — in short, that it merits further investigation (see 'The regulator', left).

Preparing for the move

These applications contain data from preclinical trials, including animal pharmacology and toxicology studies (see *Nature* 441, 544–545; 2006), manufacturing information that ensures the drug can be produced safely and consistently, and detailed clinical protocols for the proposed human studies. Problems at this stage can delay or stop the drug's move into clinical trials.

"The agencies provide guidance on how a new drug should be tested, but it's not always black and white. Interpretation makes regulatory affairs interesting and intellectually challenging," says Ann Readman, vice-president of regulatory affairs at AstraZeneca in Alderley Park, Cheshire, UK.

The IND is a pivotal point in any compound's move from lab bench to market, but it is only the start of the regulatory-affairs professional's relationship with the drug. "Once you've filed an IND, you've opened yourself up to all kinds of responsibilities," says White. Any new animal studies, or changes to the manufacturing process or clinical protocol, must be submitted to the agency. Also required are annual reports on safety information from the clinical trials, and meetings with the regulatory agencies at key points throughout the studies.

Recent safety issues, such as those involved in the

THE REGULATOR

The Office of New Drugs at the US Food and Drug Administration (FDA) receives about 1,800 Investigational New Drug applications every year, says the office's deputy director, Sandra Kweder. Regulations require that each application has to be reviewed within 30 days from the date the FDA receives it.

"When drugs come to the FDA for investigation, a number of professionals get involved in the assessment of that drug," says Kweder. Review teams made up of pharmacists, nurses, chemists, toxicologists, physicians, statisticians and microbiologists determine whether a drug can safely move forward into clinical trials. Most have doctoral degrees and some postdoctoral experience, says David Jacobson-Kram, associate director of the pharmacology and toxicology staff at the office. His department employs about 100 of the office's 800 staff.

Although salary differences can make it difficult to recruit people from industry, Jacobson-Kram believes that the learning experience offered by the Office of New Drugs is unmatched in industry.

"Often in industry you get wedded to a particular drug for your whole career, whereas at the FDA you see a whole range of drugs and different approaches to the same problems," he says. "Intellectually, it can be a whole lot more satisfying."

H.H.

IMAGES.COM/CORBIS

TGN1412 clinical trial and the lawsuits surrounding the painkiller Vioxx, have led to heightened awareness and review of the approval process. Some regulatory agencies require industry to take a proactive role and investigate potential safety issues during the product's life cycle using a pre-approved risk-management plan, increasing the amount of work for staff, says Readman.

The relationship with a drug may continue until the company stops manufacturing it. "They will work with that compound from the time it emerges from research, through development and post-approval; from birth to death of the product," says Susan Boynton, senior director of regulatory affairs at Amgen in Thousand Oaks, California.

Being the link between their company and the regulatory agency, regulatory-affairs professionals have a rare combination of skills. A science degree is usually required. "We are looking first for people who are strong scientists, people with pharmacology, life-sciences and clinical-medical backgrounds," says Skeens. According to a 2002 North American survey, about 60% of them hold at least a master's degree, and nearly 20% have a doctorate. Those with bachelor's degrees and additional training can also make the cut (see 'Learning curve', right).

But they also must have leadership skills and be impeccable oral and written communicators. That combination of skills is rare in a scientist, says Skeens. Regulatory-affairs specialists need good judgement and the confidence to make predictions about how the regulatory environment is changing. They need a knack for investigating drugs with novel mechanisms of action or new therapeutic areas, such as monoclonal antibodies, years before these will reach the market. Scientists who enjoy strategic planning, working on communications, or having a business perspective also fare well in the profession, Skeens adds.

People come from all kinds of backgrounds, says

LEARNING CURVE

Most in-house regulatory-affairs staff gain their expertise through on-the-job training, but a handful of academic programmes exist. Nestled within the top biotech and pharmaceutical hubs in the United States, these on-campus or online training programmes provide students with an introduction to the pharmaceutical, biotechnology and medical-device industries, basic regulatory requirements and an understanding of how guidelines are interpreted and enforced.

Many of the students are already working in the industry, says Larry Gundersen, director of regulatory programmes at San Diego State University.

"What we've done is try to create a product — our graduate — who has much more breadth and depth than can be achieved with three years of experience in the industry alone," he says. Since 1999, 35 students have graduated from the programme.

The Regulatory Affairs Professionals Society lists 24 undergraduate, graduate and certificate programmes available at colleges and universities around the world, including 17 in the United States and 3 in Europe.

H.H.

WEB LINK

US Food and Drug Administration

♦ www.fda.gov/cder

Regulatory Affairs Professionals Society

♦ www.raps.org/s_raps/index.asp

The Organisation for Professionals in Regulatory Affairs

♦ www.topra.org

European Medicines Agency

♦ www.emea.eu.int

Joseph Scheeren, head of global regulatory affairs at Bayer HealthCare's pharmaceuticals division in West Haven, Connecticut. Some join regulatory affairs immediately after finishing their education; others make internal moves from development or discovery, he says. New recruits can expect to receive a lot of mentoring and coaching during the first two years of their job. Regulatory professionals with experience in a specific therapeutic area are especially appealing to companies that are expanding into new areas.

Today's professionals see regulatory affairs as a career that offers continual learning and interaction with scientists and regulators. And with harmonization, changing regulations and the constant expansion of the therapeutic realm, it's never dull. "I've been in the regulatory-affairs area for almost 25 years, but I learn every day," says Scheeren.

Word of mouth

Scientists and students interested in moving into this area should contact organizations such as the Regulatory Affairs Professionals Society or the Organisation for Professionals in Regulatory Affairs. "These organizations help people get a broader understanding of what regulatory affairs is about," says Boynton, adding that people used to hear about the discipline mostly by word of mouth. "Most of us are very willing to give our time to sell our profession."

Job availability in regulatory affairs is linked to the drug-development pipeline and many companies are currently expanding the size of their regulatory-affairs departments. "We are constantly looking for high-calibre candidates," says Boynton, whose department has doubled in size in the past four or five years. "It reflects the large number of new products that we are focusing on. The pipeline is very exciting."

The regulatory-affairs profession is still young. It emerged in the 1970s and grew in popularity in the 1980s as the number of INDs skyrocketed. People's perception of it has changed radically over that time. "It's risen in its position in the company in terms of its contribution to the overall business strategy," says Boynton. Readman agrees: "In the old days, the regulatory-affairs function was seen as a negative place to be. They were the people who told you what the rules were." Today, the regulatory-affairs professional works alongside the scientist. "Their data are part of a much bigger picture," Readman says.

Hannah Hoag is a freelance writer based in Montreal, Canada.



IMAGES.COM/CORBIS

MOVERS

Paul Gilna, executive director, Community Cyber-infrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA) project, San Diego, California



2001-06: Director/manager, Center for Human Genome Studies, Los Alamos National Laboratory (LANL), New Mexico

1998-2000: Programme director, computation biology and database activity programmes, National Science Foundation, Washington DC

1994-98: Programme manager, biology and environmental research, LANL, New Mexico

A recurring theme of Paul Gilna's career has been to serve the scientific community while working at the cutting edge of computational biology. His work to set up computational tools and annotated genetic-sequence databases, such as GenBank, has aided the research of countless scientists.

After earning a degree in pharmacology at University College Dublin, Gilna pursued a PhD focused on the drugs affecting the hormones oestrogen and progesterone. Advised by mentors to enter the new world of molecular biology, he serendipitously met molecular biologist Pierre Chambon. Gilna's collaboration with Chambon's group and colleagues at the University of Chicago in Illinois culminated in the cloning and sequencing of human oestrogen and progesterone receptors.

While a postdoc at Chicago, Gilna published papers that would become the basis of current genetic testing for the predisposition to breast cancer. At this point, however, he opted out of research to use the computational-biology skills he had taught himself. A position at GenBank, the annotated collection of publicly available gene sequences at Los Alamos National Laboratory, met his goals. "I was looking for something more tangible as evidence that what I doing was of assistance to others," he says.

Gilna constructed the now familiar requirement that authors of journal articles submit gene sequences to GenBank in exchange for an accession number printed with the article. The procedure made digital data available to researchers and streamlined the process of lifting printed data from journals. He was co-principal investigator until GenBank moved to the National Institutes of Health.

After two years at the National Science Foundation, Gilna returned to Los Alamos to oversee the human-genome sequencing arm of the Department of Energy's Joint Genome Institute. One of his best achievements, he says, was aiding its transition to microbial genome sequencing.

His new role is executive director of the CAMERA project, which provides computational and data-analysis tools to decipher the collective genetic codes obtained from as-yet unculturable ocean microbes. The partnership between the J. Craig Venter Institute at the University of California, San Diego, and the San Diego Supercomputer Center will test his ability to lead disparate groups. "The challenge is not only that this is a new partnership, but that it is creating the new field of metagenomics," he says. ■
Virginia Gewin

RECRUITERS & ACADEMIA

Lessons from the jungle

Since I graduated in 2003 with a bachelor's in biology, my professional life has been an on-again, off-again drama. Should I move to a graduate programme straight away? Or should I test the waters outside of the hallowed halls of academia? I chose the latter.

I'm glad I did. I had little wind left in my jib after four tough years. The challenge has been to remain active and keep my mind astute in the ever-changing fields of biology and ecology. Small, volunteer gigs and journal-reading sufficed for a while, but I soon became more aggressive in applying for full-time positions.

I first targeted biology departments and also sought private-sector and federal positions. Many opportunities I saw were for full-time volunteers, most requiring major expense and lasting a short time. Then in early 2005 I spotted a volunteer post at the Primate Habituation Programme at the Dzanga-Ndoki National Park in the Central African Republic. I had never been there and had never worked with primates. With a one-year commitment, I wouldn't feel like a tourist. I applied, got the position and was soon on my way.

The goal of the programme, based at a camp called Bai Hokou, is to acclimatize western lowland gorillas and semi-terrestrial mangabeys to the

presence of humans. An eco-tourism programme was set up with the backing of the government, the conservation charity WWF and the German aid agency GTZ. The revenue is placed in a programme for local communities. Tourists have been visiting since 2001, although acclimatization efforts are labour-intensive and ongoing.

The position enabled me to witness a new model of wildlife conservation that had to weather poachers and limited government supervision. It was a stark contrast to the carefully choreographed efforts I'd seen in the United States.

Because I was either fixing fuses, cooking over a wood fire or following gorillas, I gave little thought to what would come next. So, since coming home, I'm once again excruciatingly over-analysing my career options.

I now plan to capitalize on the discipline and intellectual stimulation I gained in the Congo basin by pursuing a master's degree in evolutionary biology with an emphasis on conservation. In the mean time, I'm doing public presentations about working with wild animals in an effort to raise awareness of conservation issues. ■

Ayres Christ is currently splitting time between Washington DC and Tennessee.

GRADUATE JOURNAL

Write and wrong

There are many books on scientific methodology, but guidelines for writing a good scientific paper are harder to come by. My month-long battle with a manuscript has made me ponder many aspects of writing.

First, there is the choice of scope. The best articles put their subject in a wider framework, but it's a fine line between relevant and ridiculous parallels. As a sociobiologist, I'm wary of making careless comparisons of ants to humans, despite frequent requests from friends both within and outside science.

Then there is the eternal temptation to promise more than your results deliver. My paper is on kin selection and insects. Do I suggest that it will change the way we study ants, the way we see social insects, or the roots of social interaction? The insanely inflated claims of some papers make me side with a friend of mine, who once suggested that articles should include only methods and results, leaving interpretation to the reader.

And how much room is there for speculation? Writing my paper has inspired many ideas about my ant system that remain untested. Gratuitous musings may transmit the ideas, but they make many reviewers see red.

After weeks of rewrites, I finally submitted my manuscript. At the moment, I have no clear picture of what it says or, indeed, why any of it matters. But I'm hopeful that given some time, I will be pleasantly surprised by my handiwork. ■
Katja Bargum is a graduate student in the Department of Biological and Environmental Sciences at the University of Helsinki, Finland.

Check elastic before jumping

Paralysis of the senses, temporarily.

Neal Asher

Standing before the museum display case, I listened to the cracking of stun rifles outside as CCM — Combined Corporate Military — engaged with the few government troops who could be bothered to fight.

"It's a crappy world," said someone standing nearby.

"Oh really," I said, peering into the case, and remembering my recent hospital visit, which had occurred just a few days before a crowd of some 500,000 drove everyone out of the EU Parliament buildings and then burnt them to the ground...

It felt like I was using the scroll wheel on an old-style computer mouse to work my way down the list of news items, then it felt like I clicked the mouse button to select a news item — National Health Service software having some way to go to catch up with the new interactive displays and virtualities. Of course, I hadn't moved a finger since that moment I decided to flout the EU-wide health-and-safety ban on urban bungee jumping and miscalculated the length of my elastic. My shattered skull and snapped spine had left me unable to do more than blink and swallow. A temporary cortical implant, fitted just before they epoxied my skull back together, provided the sensation of movement from my paralysed limbs and enabled me to use the entertainment system here.

The news item played out in an 'Infinity Deepscreen' holographic display seemingly hanging in mid-air over my bed — this illusion provided by the multicoloured lasers, which were tracking my eyes from my comm unit, at present stuck to the front of my monitor suit. Apparently, Allied Energy was drilling for the carbon dioxide that had been pumped down into the gas and oil fields at the beginning of this century. The hope being that those megatonnes of gas released into the atmosphere will raise temperatures high enough to offset the coming Ice Age by another century. The Green Socialists, still retaining a majority in the European Parliament, were pushing for more stringent controls on energy usage, their contention being that the numerous wind-power projects, having upset the weather patterns of the planet, need to be closed down. I sighed — I could manage that — and was about to move on to the next news item when my consultant interrupted the display.

"How are you today, Joe?"

"Getting a little bored with the primitive entertainment here and rather concerned that you don't seem to know my name." I didn't actually speak the words, since they were a product of my implant and a voice synthesizer, which made me sound like Humphrey Bogart — the new virtual screen idol.

In the display my consultant peered at his paper notes. "Okay, Joe, I'll give it to you straight." It must have been some sort of glitch in the system, since I'd commed others in the hospital and apparently this consultant program called all his patients



Joe. "You're in here for the long haul. The new nerve tissue is growing well, but without invasive surgery, it will take some time for the breaks in your vertebrae to knit. You will have to spend a year in a motorized exoskeleton during..."

"What about CT nerve-stimulus and keyhole bone welding?" I interrupted.

"NHS funding," he supplied.

"What? Things weren't this bad after I protested the HSE ban on hang-gliding, and that was only two months ago."

"Ah yes, I see," he said, once again peering at my notes. He froze for a moment, which seemed to happen whenever he updated, then continued in a slightly different tone. "I see that then you were a European citizen, but now it appears you are a citizen of Ubermart?"

"Yes, I certainly am — I'd rather my taxes went to a more deserving cause." Really, what had finally persuaded me to sign up

for corporate citizenship was the last burglary of my apartment. The guy breaking in managed to stab himself with the screwdriver he tried to jemmy the door with and I called him an ambulance. The police arrested me for negligence and then helped the burglar to sue me for damages. Apparently he was 'disadvantaged', unable to read or write due to ritalin abuse and dietary cerebral damage. Corporates tend not to suffer the same problems. Corporate police tend to take an old-fashioned approach to policing, usually in CCTV blackspots.

"If you could confirm your Ubermart citizenship number?"

I recited it for him quickly and his image froze again. When he thawed out he said, "Thank you citizen Philips. Do you mind if I call you Derek?"

"Of course not, it's my name."

"Well, Derek, I am sure we can improve on my previous prognosis. Before we get to that, would you like to enjoy our virtual-reality facilities during your stay?"

"How long will that be?"

"Oh, I think we can have you out of here by this afternoon, Derek."

"That's okay — I'll stick with what I have for the moment." I'd forgotten all about Ubermart's health plan, which as one of their citizens I could also enjoy.

The rest of the day just seemed to speed past. Actual human nurses came in with a suit hoist, asked me kindly if I was ready while simultaneously instructing my suit to inject a really trippy narcotic. Just before I went under I observed a surgical robot cracking open my monitor suit then poisoning over me while extending its glittering cutlery.

I awoke amid clean sheets. Walked out of the hospital only an hour later.

The sounds of stun rifles firing seemed to be decreasing.

"You don't agree that it's a crappy world?" asked the guy standing next to me.

"Okay, it's crappy, but it could be worse."

He harrumphed and moved away. I returned my attention to the display case and smilingly inspected the wheelchair resting there, then the dates on the plaque indicating when the first one was invented, and when the last one was used.

■ **Neal Asher is now on his third three-book contract with Macmillan, is translated into nine languages and reckons he's doing okay. His latest book, *Polity Agent*, will be available at the end of this year, <http://freespace.virgin.net/n.asher>.**

JACEY